University of Utah

Performance-Based Compensation Pilot Program Evaluation

Evaluation Technical Report

Andrea K. Rorrer, Ph.D., Kristin L. Swenson, Ph.D., Randy Raphael

THE UNIVERSITY OF UTAH
UTAH EDUCATION POLICY CENTER

November 2011

EXECUTIVE SUMMARY

The Performance-based Compensation Pilot Program was created in 2009 by the Utah State Legislature (HB 328, codified as 53A-17A-163). The intent of the program is suggested by a separate resolution introduced in the same year (HJR 13), which called for teacher compensation systems that would support student achievement and quality instruction, be adequately funded, promote collaboration, be flexible in responding to local needs, use fair criteria, be open to all eligible parties, and align with existing programs and school improvement plans.

Background

Program Components. The program was designed to include performance in three areas: student learning gains (40%), instructional quality (40%) and parent satisfaction (20%).

Selection of Pilot Schools. In Spring 2009, Utah elementary schools were invited to apply for the program (State Board of Education R277-113). Out of 19 schools that applied, five schools—three district and two charter schools—were selected to participate. Collectively, the pilot schools were significantly different from other schools in the State on most student demographic and academic variables. For example, pilot schools had less racial diversity, a smaller concentration of students eligible for Free and Reduced Lunch, and students with higher average test scores. As a result, these differences limit the generalizability of the findings to other schools.

Evaluation Methods

The evaluation employed a mixed-methods approach that addressed how the compensation-based plans were developed and implemented and the associated student and teacher outcomes. The implementation process was primarily assessed through focus groups and individual interviews with 115 district administrators, principals, teachers, and parents across the schools, and open-ended responses to teacher surveys. Impact was primarily assessed through statistical analysis of student test scores and closed-ended responses to teacher and parent surveys. *It is imperative to note that individual school plans did not necessarily include analysis of CRT data in determining bonuses*. However, due to the variability in the school plans, this evaluation does not attempt to evaluate the level of fidelity of implementation to the accepted plans, as they were proposed.

Key Findings

The design of the program permitted each school to establish its own process, tailor its plan to its *unique* situation, and utilize different measures of the three components. In fact, each school did develop a unique approach. This contributed greatly to faculty ownership of the program, but also made it impossible to compare the effectiveness of plans or evaluate the impact across the pilot schools using the plans to establish comparable measures.

Planning. Although planning roles and teacher input varied across schools, 80% of all teachers reported being involved with the planning process. Plans were typically aligned with programs that already existed in the schools, which reduced the burden of implementation on teachers. However, the lack of technical assistance made the planning process more difficult. All plans included individual measures to ensure differentiation of bonus pay (as required), and some plans incorporated a tiered approach that provided a portion of the bonuses on the basis of team and/or school-wide goals. All plans promoted collaboration

over competition. Interestingly, no school thought that its plan could be replicated in its entirety by another school.

Implementation. Principals played a key role in implementation, and both principals and teachers found that the process of collaboratively planning and implementing the program brought about positive changes in their professional practice. Teachers expressed appreciation of the bonuses received as recognition for a job well done. The average bonus paid to a teacher during the implementation year was $1786. When asked whether the rewards offered were of sufficient value, only half of teachers (51%) agreed that they were. There was uncertainty as to whether the size of bonuses in the pilot would be sufficient to motivate action in the long-run and whether positive changes introduced during the program would be able to continue in the absence of continued funding. From interviews and focus groups, it was apparent that a number of teachers saw the requirements of the program, however fair and transparent, as taking away from their teaching responsibilities.

Student Achievement. *Except for the schools that used Criterion Referenced Test (CRT) as a measurement, school participation in the program was not associated with changes in average CRT scores.* However, schools did report improvement on the measures that they included in their plans. Schools also increased the formative use of data. Although not all teachers perceived the program as being related to improvement, 51% of teachers agreed that the program improved the learning experience for students.

Instructional Quality. Peer and principal observation of classroom teaching as well as self-reflection were among the most commonly used measures of instructional quality. Both teachers and principals reported that building mutual trust was essential in obtaining the intended professional development benefits of classroom observation. Of the teachers who taught math or language arts, 56% of them said that they modified their instruction for one or both of these topics as a result of the program. Of the teachers that did modify their instruction, 74% agreed that the program had a positive impact on students. Just less than half (48%) of teachers agreed that the program improved the school experience for teachers.

Parent Satisfaction. Teachers agreed that they had increased communication with parents as a result of this program. However, teachers ranked parent satisfaction as the least important factor on which to base performance-based compensation. Virtually all parents (99%) who responded to the survey reported at least quarterly contact with their child's teacher, and two-thirds (68%) reported at least monthly contact. Parents gave equally high grades to both schools and teachers. A factor analysis showed satisfaction with the teacher and satisfaction with the school to be distinct factors. Approximately 44% of the parents who responded to the survey knew about performance-based compensation program. Of the parents who were aware of the program, 44% (approximately 20% of all parents who responded) gave positive comments when asked about the impact on their child's learning.

CONSIDERATIONS

The five schools that participated in Utah's Performance-Based Compensation Program—Ashman, Canyon Rim, Manila, Midway, and Wasatch Peak Elementary Schools—were intricately involved throughout the planning and implementation process. From their experience and the analysis of the implementation process and associated outcomes, we provide the following considerations for future planning and development of a strategic compensation program.

Pre-Implementation

- Plan for sustainability (e.g. sufficient resources, including amount and longevity), including allocation of funding for multiple years to support a planning year and at least three years of implementation.
- Provide guidance for implementation, including general framework (e.g., instructional quality, student achievement, and student, parent, and community satisfaction) for the program.
- Consider a menu of potential measures for schools and districts to select from for instructional quality, student achievement, and student, parent, and community satisfaction.
- Ensure availability of program to diverse schools.
- Establish a network of program participants with regular communication, dissemination of information and success, and opportunities to convene participants, including administrators and teachers.

School and District Planning

- Provide leadership for the vision for school and definition of the purpose of the school's involvement in the initiative.
- Ensure and/or develop a climate of trust and collaboration.
- Develop a planning team and process that is inclusive and representative of the school staff (e.g., teachers, specialists, paraprofessionals) and community (e.g., parents, school community council members, district administrators).
- Provide compensation for those engaged in development of plan.
- Support for the development of multiple measures appropriate for school context within the instructional quality, student achievement, and student, parent, and community satisfaction.
- Use data to inform selection of measures and indicators for success.
- Provide technical assistance for the development and use of the common assessment tools.
- Develop tiered plan that acknowledge the individual's contribution, a team's contributions, the overall school's progress.
- Utilize external resources (e.g., district and state staff, consultants) to support planning process and development of plan.
- Plan for alignment with other key initiatives.
- Allow schools to create plans that are customized to the individual school but require at least one common assessment for each of the areas of evaluation.
- Communicate consistently with all stakeholders regarding process, timelines, expectations, and anticipated outcomes.
- Expend time and effort to develop buy-in among all stakeholders in the process.
- Use all measures formatively.

Implementation

- Maintain communication consistently with all stakeholders regarding process, timelines, expectations, and anticipated outcomes.
- Focus on instruction and instructional quality.
- Provide support for school leadership to administer the program.
- Use data to utilize the lessons learned from the implementation year.
- Build capacity (e.g., professional development, training, shared learning, networks) to develop, implement, and refine plans.
- Provide professional development to support staff (e.g., improving instructional quality, increasing student achievement, improve student and/or parent satisfaction, development of

common and/or formative assessments, use of assessment data, data analysis, structured improvement based on data).

- Monitor implementation for plan improvement and fair application of the plan.
- Utilize annual results for school improvement efforts.

ACKNOWLEDGEMENTS

The Utah Education Policy Center would like to thank the schools and districts, including the administrators, teachers, and parents for their assistance in facilitating data collection and providing feedback about the performance-based compensation pilot program. Their time and participation in this evaluation are greatly valued. We also appreciate the Utah State Office of Education for their guidance in the development of the evaluation.

CONTENTS

Tables

Figures

EVALUATION FOCUS

The Performance-based Compensation Pilot Program was created by the Utah State Legislature in 2009. The house bill that created the pilot (§ 53A-17A-163, 2009 HB 328 sponsored by Representative Gregory Hughes and Senator Howard Stephenson R-Draper) provided for the development and implementation of the initiative for five elementary schools. Applicants for the program had to address how the school's participation would support student achievement and quality instruction, be adequately funded, promote collaboration, be flexible in responding to local needs, use fair criteria, be open to all eligible parties, and align with existing programs and school improvement plans. The Utah Education Policy Center (UEPC) was retained by the Utah State Office of Education (USOE) to conduct an evaluation of the implementation of the Performance-based Compensation Programs. The goals of this evaluation were to determine how the development and implementation of the plans aligned with the resolutions of the program and to evaluate the impact that the program had on professional practice and students. This evaluation does not attempt to evaluate the level of fidelity of implementation to the accepted plans, as they were proposed.

BACKGROUND AND CONTEXT

Student Characteristics

Traditional and Charter Elementary schools in Utah were invited to apply for the Performance-based Compensation Pilot Program towards the end of the 2008-2009 school year. From an application pool of 19, five schools, three traditional and two charter, were selected to participate in the program. The selected schools were significantly different from one another on most student demographic and academic variables. The students from the performance-based compensation schools, collectively, were also different from the other elementary school students in the state. Below is a brief overview of the student characteristics for each of the pilot schools. In addition, this information is presented for comparison in Table 1.

- **Ashman Elementary** is a rural K-2 Elementary school in Sevier County. As of October 2010, Ashman had 514 students with approximately 6% minority students and approximately 53% of students receiving free or reduced lunch. The average class size at Ashman was about 21 students per class.

- **Canyon Rim Academy** is a K-6 Charter school in the Millcreek area of Salt Lake City that was previously an elementary school in Granite School District. As of October 2010, Canyon Rim had 530 students with approximately 15% minority students and approximately 20% of students receiving free or reduced lunch. The average class size at Canyon rim was about 24 students per class.

- **Manila Elementary** school is a K-6 Elementary school in Alpine School District. As of October 2010, Manila had 820 students with approximately 3% minority students and approximately 17% of students receiving free or reduced lunch. The average class size at Manila was about 29 students per class.

- **Midway Elementary** is a K-4 Elementary school in Wasatch School District. As of October 2010, Midway had 509 students with approximately 18% minority students and approximately 37% of students receiving free or reduced lunch. The average class size at Midway was 25 students per class.

- **Wasatch Peak Academy** is a K-6 Charter school in North Salt Lake that opened in 2005. As of October 2010 Wasatch Peak had 375 students with approximately 21% minority students and approximately 21% of students receiving free or reduced lunch. The average class size at Wasatch Peak was 24 students per class.

Table 1. Demographic Characteristics of Each Pilot School

School	School Type	Grades Served	Enrollment	Percent Students of Color	Percent Students Receiving Free or Reduced Lunch	Average Class Size
Ashman	Traditional	K-2	514	6%	53%	21
Canyon Rim	Charter	K-6	530	15%	20%	24
Manila	Traditional	K-6	820	3%	17%	29
Midway	Traditional	K-4	509	18%	37%	25
Wasatch Peak	Charter	K-6	375	21%	21%	24

In Utah, approximately 22% of all elementary school students are students of color and approximately 39% of all elementary students receive free or reduced lunch. The average class size in Utah elementary schools is about 24 students per class. When compared with these parameters, the schools that participated in the Performance-based Compensation Pilot were less racially diverse and more financially advantaged than the average students in the state, with approximately the same average class sizes. Participating schools had higher Criterion Referenced Test (CRT) scores in the 2009-10 school year, which was one year prior to the implementation of the program, than other schools in the state for language arts, math, and science.

Figure 1 shows the differences between Performance-based Compensation and non- Performance-based Compensation schools on the demographic variables and Figure 2 shows the differences between performance-based compensation and non- Performance-based Compensation schools on the academic variables.

Figure 1. Demographic Characteristics of Performance-based Compensation and Non- Performance-based Compensation Elementary Schools

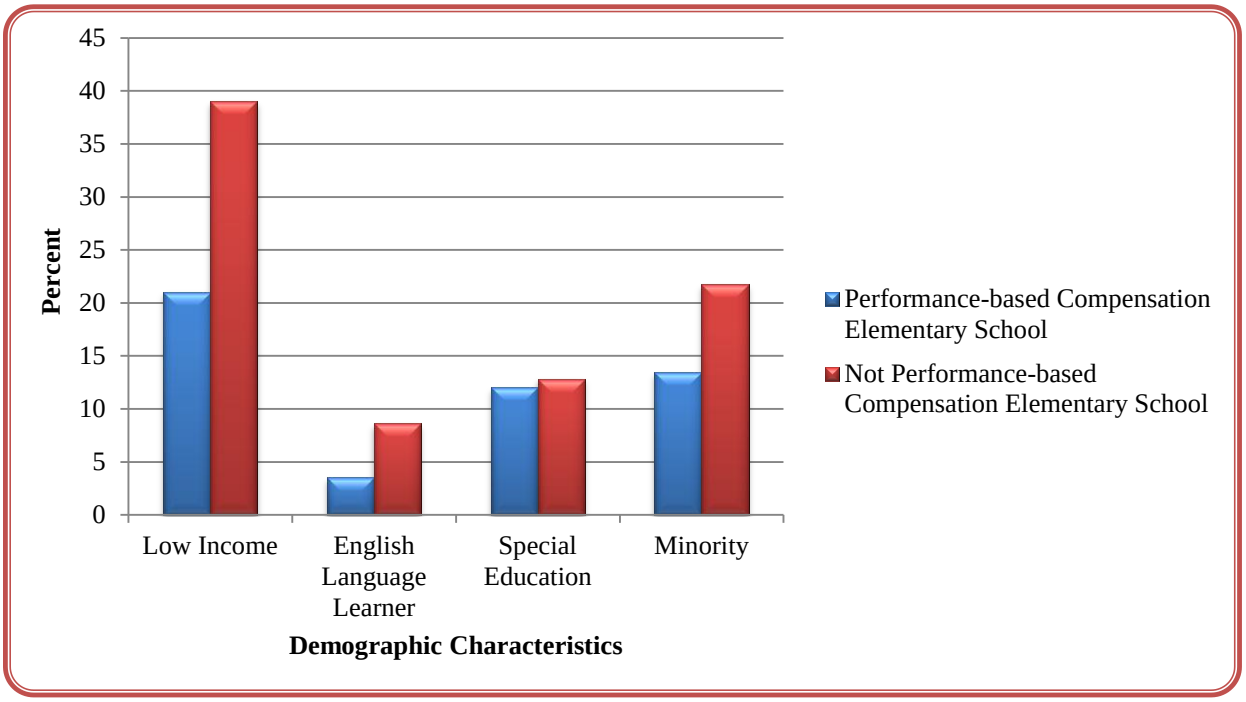
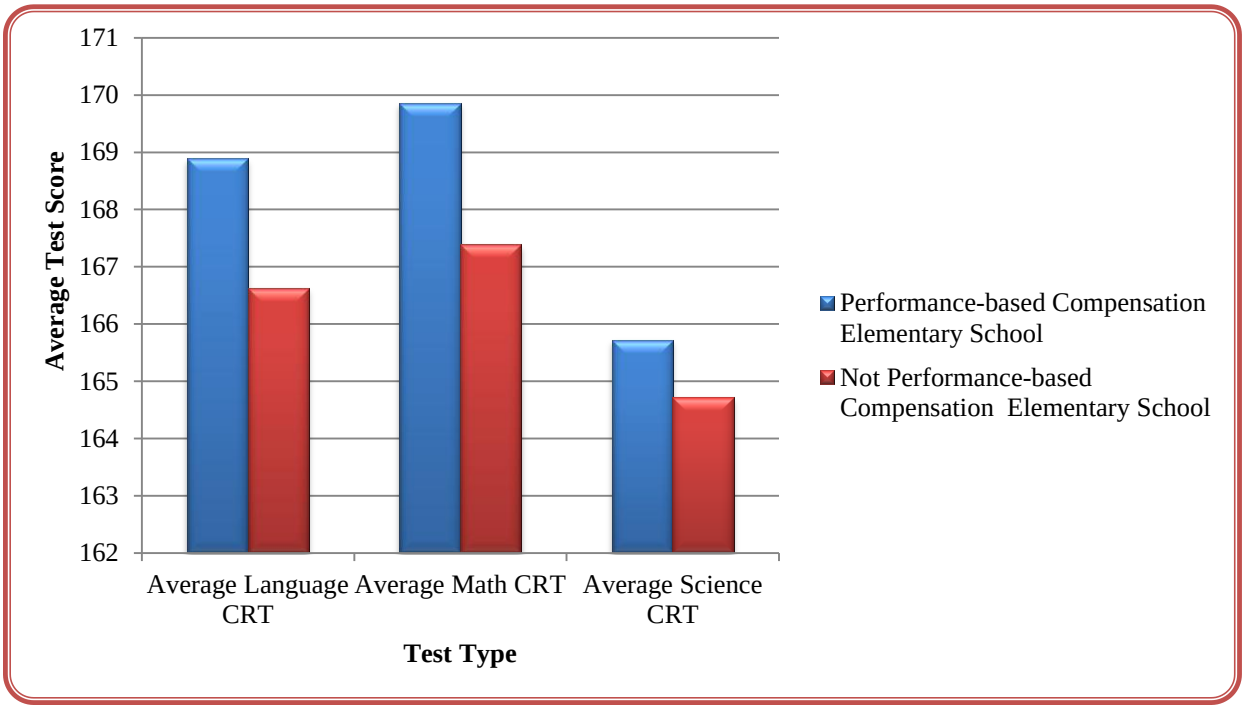


Figure 2. Academic Characteristics of Performance-based Compensation and Non- Performance-based Compensation Elementary Schools.



Faculty Characteristics

Table 2 profiles the faculty of each participating school in comparison to faculty at regular elementary schools in the state as a whole. In general, the faculties of pilot schools were smaller in size (total number of licensed faculty averaged 34 compared to 40 for the state), younger in age (on average 39 years old compared to 43 for the state), and had a higher proportion of women (approximately 95% compared to 90% for the state). With respect to qualifications, faculty at pilot schools have less experience on average (8 years overall and 5 in teaching specifically, compared to 10 and 8, respectively, for the state) and lower educational attainment (54% with a graduate degree, compared to 62% for the state). However, faculty performance on licensure tests was equal to the state and superior to the nation (*Content Knowledge* median scale score for pilot schools = 169, state = 170, nation = 164; *Principles of Learning and Teaching* median scale score for pilot schools = 178, state = 178, nation = 175). Finally, the schools varied considerably in their use of instructional support staff (from 12% to 48% of total FTEs compared to 28% for the state; most instructional support staff are paraprofessionals).

Table 2. Faculty Profiles of Participating Schools and State, 2010-11

	Public Regular Elementary Schools	District Schools			Charter Schools	
		Ashman	Manila	Midway	Canyon Rim Academy	Wasatch Peak Academy
Licensed Faculty per school	**40**	25	36	36	34	37
DEMOGRAPHY						
Percent Female	**89.90%**	92.00%	92.30%	94.40%	97.10%	94.90%
Average Age (as of July 1, 2010)	**43.1**	46.4	37.1	41	38.9	34.9
Race						
White	**86.90%**	100.00%	94.90%	94.40%	68.60%	64.10%
Minority	**4.20%**	0.00%	0.00%	2.80%	2.90%	28.20%
Unknown	**8.90%**	0.00%	5.10%	2.80%	28.60%	7.70%
QUALIFICATIONS						
Years of Experience in Education						
In Total Average	**10.3**	12.4	8.6	11.1	7.3	4.1
In Regular Elementary Teaching	**7.4**	7.8	6.5	6	5	2.2
Highest Educational Attainment						
Master's or Higher	**33.40%**	76.00%	28.20%	63.90%	34.30%	23.10%
Praxis Scale Score						
Content Knowledge Median (National = 164)	**170**	158	173	169	170	165
Median Principles of Teaching and Learning (National = 175)	**178**	173	178	178	181	179
ASSIGNMENTS						
Licensed Faculty per school (FTEs)	**38.207**	25	35.44	34.5	29.281	32.39

Classroom Instruction	Public Regular Elementary Schools	District Schools			Charter Schools	
		Ashman	Manila	Midway	Canyon Rim Academy	Wasatch Peak Academy
Regular Classroom Teachers	**65.10%**	80.00%	86.50%	65.20%	68.30%	47.90%
Special Education Teachers	**7.00%**	8.00%	2.80%	11.60%	3.40%	4.60%
Instructional Support Staff	**27.80%**	12.00%	10.60%	23.20%	28.20%	47.50%

Note. The basic data source for this and other tables regarding teacher characteristics and salary is the USOE CACTUS (educator licensing) database, which is available to the UEPC as a founding partner with the USOE and other Utah state agencies in the Utah Data Alliance. The CACTUS Database is described at http://www.schools.utah.gov/cactus. Persons included were limited to those who had an assignment requiring a license issued by Utah State Board of Education, as CACTUS does not include classified (non-licensed) staff. Regarding educational attainment, paraprofessionals are not required to hold a bachelor's degree and, for some persons, degree data is missing from CACTUS.

Retention

One of the most important outcome measures of performance-based compensation evaluations in the economic literature is retention of staff. There is no way to properly assess the impact on retention after only one year of implementation. As such, the retention patterns of each of the pilot schools are described here to provide context for the program. Table 3 - Table 7 show the history of faculty size and retention for each of the five participating schools. The two key columns for interpretation are the "turnover rate," indicating licensed faculty who left by the end of the preceding year, and the column furthest to the right, which indicates the percentage of licensed faculty who are new to the school in that year. There was considerable variation in turnover across schools and within schools over the years, no doubt depending on fluctuations in student population as well as the working conditions of the school, among other factors. The median turnover rate ranged from 13% to 25% and the "new (at school) in year" rate ranged from 10% to 28%. The two charter schools exhibited more movement of faculty in and out.

Table 3. Faculty Size and Retention: Ashman Elementary School in Sevier School District, 1997-98 to 2010-11

Ashman Elementary					New at School	
School Year	Head Count	Replace-ments	Turnover Rate	New Staff	Count	As % of Staff
2000-01	30	5	16%	-2	3	10%
2001-02	32	3	10%	2	5	16%
2002-03	37	4	13%	5	9	24%
2003-04	37	5	14%	0	5	14%
2004-05	37	5	14%	0	5	14%
2005-06	35	4	11%	-2	2	6%
2006-07	27	10	29%	-8	2	7%
2007-08	28	1	4%	1	2	7%
2008-09	27	5	18%	-1	4	15%
2009-10	25	3	11%	-2	1	4%
2010-11	25	4	16%	0	4	16%

Note. The dotted line indicates the beginning of the Performance-based Compensation pilot program. The analysis uses head count instead of full time equivalent (FTE), which is arguably more accurate in reflecting expansion or retraction of the faculty, but (a) it turns out that the two measures exhibit highly similar trends (r = .98, total years n = 91), (b) the decision to stay, leave

or join is a phenomenon that pertains to the individual as whole, (c) "replacement" is more difficult to operationalize in terms of FTE, and (d) change in FTEs can be an administrative response to as well as a driver of the decision to leave.

Table 4. Faculty Size and Retention: Canyon Rim Academy, Charter Elementary School in Salt Lake City, 2007-08 to 2010-11

Canyon Rim Academy					New at School	
School Year	Head Count	Replace-ments	Turnover Rate	New Staff	Count	As % of Staff
2007-08	29					
2008-09	29	8	28%	0	8	28%
2009-10	31	4	14%	2	6	19%
2010-11	34	6	19%	3	9	26%

See the note to Table 3.

Table 5. Faculty Size and Retention: Manila Elementary School in Alpine School District, 1997-98 to 2010-11

Manila Elementary					New at School	
School Year	Head Count	Replace-ments	Turnover Rate	New Staff	Count	As % of Staff
2000-01	39	22	42%	-14	8	21%
2001-02	38	12	31%	-1	11	29%
2002-03	34	11	29%	-4	7	21%
2003-04	37	1	3%	3	4	11%
2004-05	38	8	22%	1	9	24%
2005-06	39	10	26%	1	11	28%
2006-07	38	12	31%	-1	11	29%
2007-08	39	6	16%	1	7	18%
2008-09	40	12	31%	1	13	33%
2009-10	40	7	18%	0	7	18%
2010-11	36	10	25%	-4	6	17%

See the note to Table 3.

Table 6. Faculty Size and Retention: Midway Elementary School in Wasatch School District, 1997-98 to 2010-11

Midway Elementary					New at School	
School Year	Head Count	Replace-ments	Turnover Rate	New Staff	Count	As % of Staff
2000-01	32	3	12%	6	9	28%
2001-02	38	3	9%	6	9	24%
2002-03	36	6	16%	-2	4	11%
2003-04	43	6	17%	7	13	30%
2004-05	48	5	12%	5	10	21%
2005-06	50	8	17%	2	10	20%
2006-07	38	22	44%	-12	10	26%
2007-08	40	3	8%	2	5	13%
2008-09	45	7	18%	5	12	27%
2009-10	36	13	29%	-9	4	11%
2010-11	36	3	8%	0	3	8%

See the note to Table 3.

Table 7. Faculty Size and Retention: Wasatch Peak Academy, Charter Elementary School in North Salt Lake City, 2005-06 to 2010-11

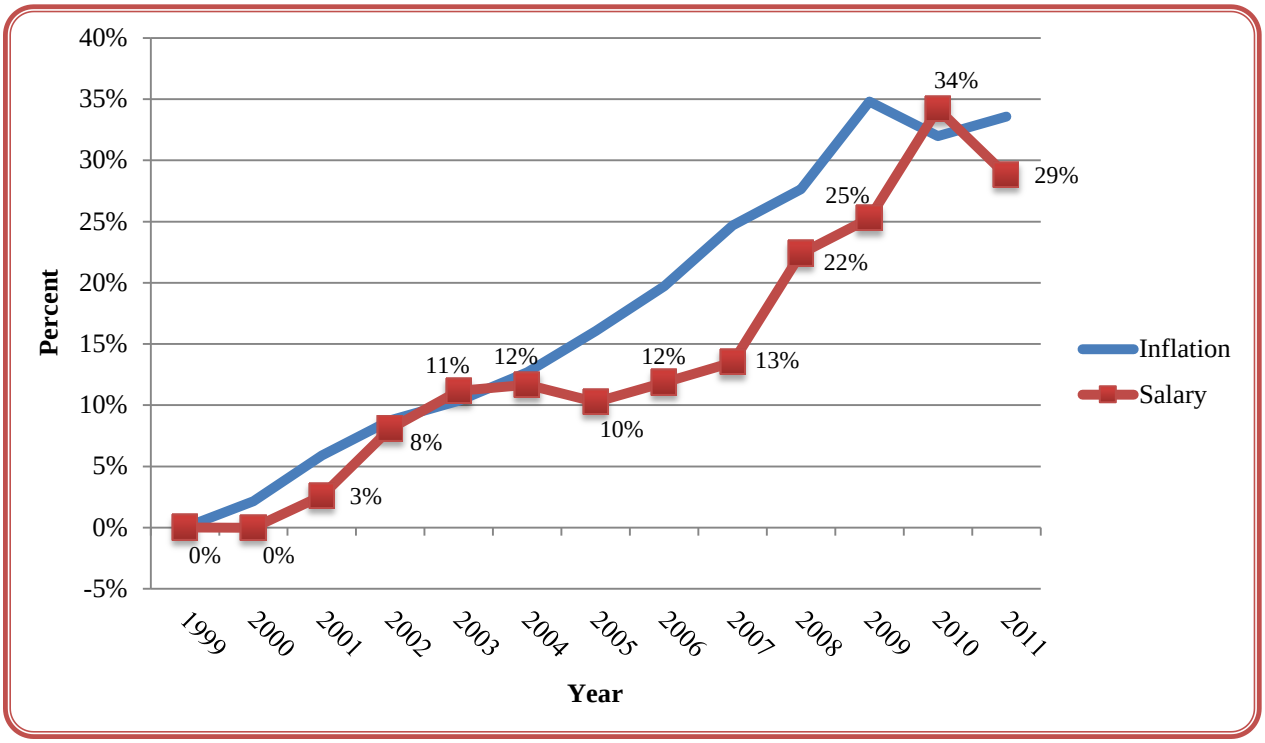
Wasatch Peak Academy					New at School	
School Year	Head Count	Replace-ments	Turnover Rate	New Staff	Count	as Pct. of Staff
2005-06	17					
2006-07	28	12	71%	11	23	82%
2007-08	24	9	32%	-4	5	21%
2008-09	33	6	25%	9	15	45%
2009-10	36	7	21%	3	10	28%
2010-11	37	7	19%	1	8	22%

See the note to Table 3.

Teacher Salaries

Figure 3 shows growth in average teacher salary in comparison to inflation for all elementary regular education classroom teachers (those likely to be teachers of record or primarily responsible for the performance of students on tests of academic achievement) in Utah public schools for the period for which sufficient salary data are available. In general, teachers have seen real growth in their salaries over the base year — they made almost $10,000 more dollars in 2011 than in 1999 (for an average yearly raise of about $760) with a decline from 2010 to 2011. Importantly, in most years growth in salary was not sufficient to keep pace with inflation.

Figure 3. Cumulative Real Growth in Utah K-6 Public School Teacher Salary and Inflation, Each Year 1999-2000 to 2010-11 Compared to Base Year 1998-99

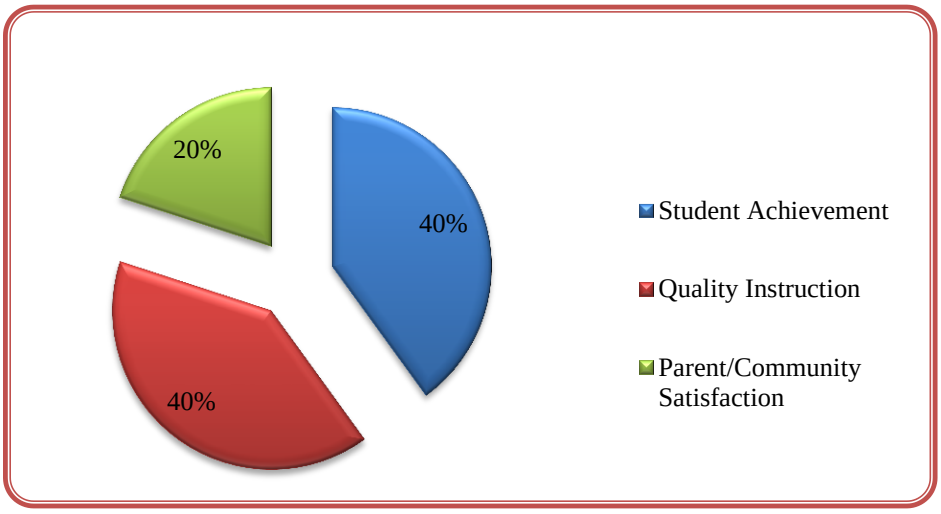


Note. 1998-99 is the earliest school year for which sufficient educator salary data is available to support analysis. Inflation is based on the CPI-U (http://www.bls.gov/cpi/cpid1108.pdf, Table 24). Salaries used were limited to those of faculty who met the following six conditions: Staff Category = 701 (Kindergarten) or 710 (Elementary); Total FTE = 1.000, Intern = No, Contract Salary ≥ 10,400 (modal value of Contract Hours times current Utah minimum wage or 1440 hours * $7.25); Private or Federal School = 0; Begin (Hire) Data < October.

Performance-based Compensation Plans

After selection for inclusion the Performance-based Compensation Program, schools were permitted to develop their own performance compensation plan, as long as the plan adhered to some common parameters, including 40% for quality instruction, 40% for student outcome, and 20% for parent satisfaction. For example, the proportions of funding for each program component—student outcome, quality instruction, and parent satisfaction are presented in Figure 4.

Figure 4. Parameters for Performance-based Compensation



The school plans were also required to stipulate how teacher pay would be differentiated (i.e., not all teachers could make the same amount). The schools were allotted $2000 per faculty member for the plan development year. A brief overview of the measures in each plan submitted by the schools at the conclusion of the planning year is included in Table 8. More detailed descriptions of each of the plans follow.

Table 8. Performance Pay Measures per School Plan

School	Quality Instruction	Student Achievement	Parent/Community Satisfaction
Ashman	•Art and Science of Teaching Class •iObservation (goals and observations)	•Literacy: NWEA Primary Map (K-1) •Literacy: NWEA Map (2) (both were LA measures)	•Parent Survey
Canyon Rim	•UPTS standard 4--five goals •UPTS standards 1-3--2 or 3 goals	•Utah Core Math Standards	•Parent Survey
Manila	•Alpine Collaborative incentive plan •Manila Observation Checklist •Manila Quality Teaching Goals Rubric	•LA and Math CRT (2-6) •Utah Kindergarten Assessment (K) •District Math and LA Assessment (1) •Direct Reading Assessment (K-6) •Common Assessment of Essential Standards (K-6)	•Parent Survey •Survey of Parents at Collaborative Presentation
Midway	•PLC Requirements •Mentorship Requirements	•AYP •District created benchmark test in math	•Parent Survey •Parent Communication Log •Participation in PTA events

School	Quality Instruction	Student Achievement	Parent/Community Satisfaction
Wasatch Peak	•Checklist of Essential Components and Conditions for Effective Reading Instruction •UPTS rubric •Self Reflection based on UPTS	•AYP (language) •CRT (3-6) or DIBELS (K-2) •Case studies and goals	•Parent Survey (Wasatch Peak)

Note. Acronyms used in this table: Adequate Yearly Progress (AYP), Language Arts (LA), Northwest Evaluation Association (NWEA), Professional Learning Community (PLC), and Parent Teacher Association (PTA), Utah Professional Teaching Standards

Ashman

The Ashman Elementary School plan was primarily developed by the principal rather than by the collaborative faculty as was done in the other schools. This decision was made to reduce the burden placed on teachers who were already dealing with significant physical changes in the school.

Quality Instruction

The Ashman plan included the use of the iObservation system to evaluate quality of instruction. Prior to beginning the quality of instruction funding plan, teachers had to qualify for participation by completing a study course of *The Art and Science of Teaching*. After completion of the course, teachers developed professional improvement plans by selecting at least 3 of the 41 strategies on the iObservation protocol. Once the strategies were selected, teachers were evaluated on their implementation of the strategies by the principal (five times), by the literacy coach (five times), by self-reflection based on video tapes (five times) and by peer reviews (two times). This schedule required 14 observations of the teacher over the course of the school year. Teachers were awarded if they were observed all 14 times and the observations showed evidence of improvement over time. The awarding of performance pay was tied to a "growth trend line" that either went up ($800), was static ($400), or went down ($200).

Student Performance

Ashman had focused on literacy for a number of years so it continued to use literacy benchmarks as indicators of academic success. The assessments used in this endeavor were the North West Evaluation Association (NWEA) primary map (for K-1 students) and the NWEA Map for students in the second grade. These assessment tools were strongly endorsed by the teachers, with 92% of teachers expressing confidence in the measures. Student achievement was to be assessed at the teacher, grade, and school level. The growth scoring on the NWEA takes into account individual differences of each child in determining growth. Instruction resulting in student growth between 0 and 50% was considered ineffective, between 50% and 100% was considered developing, and between 100% and 150% was considered effective and greater than 150% was considered highly effective. Pay for student performance was based on these categories.

Parent Satisfaction

A parent satisfaction survey was constructed using the KIVA process with the guidance of the Utah Education Policy Center. These surveys were intended initially to be sent out twice during the school year and teachers who had at least 80% of questions answered with a satisfactory or higher collected $100. This payment of $100 per survey session or $200 per year is only 10% of the maximum of $2000 available to each teacher. It is unclear from reading the plan if the other $200 was awarded or on what basis it was awarded.

Canyon Rim

Quality Instruction

Quality of instruction at Canyon Rim was assessed in two different ways, both of which focused on Utah Professional Teacher Standards (UPTS). The first measurement focused only on the fourth goal of the UPTS: assessing and evaluating student learning. To earn the bonus pay for this measure, teachers were required to advance to proficiency, advance from proficiency, or maintain master status on each of the five goals. Each goal was treated independently and teachers received the pay multiplied by the number of goals that they achieved. The second measurement allowed the teachers to select two or three goals from the 1st, 2nd, and 3rd standards. The plan stated that all teachers who accomplished their goals were to receive an equal portion of the money. There is no distinction made between teachers with two and teachers with three goals so an all or nothing standard is assumed (i.e. teachers either did or did not meet all goals). An administrator was to meet with the teachers in March to go over the UPTS rubric and determine if goals had or had not been met.

Student Performance

Student performance at Canyon Rim in math was to be assessed at each grade level with all teachers setting grade level goals. The plan states that the Denver Public Schools Student Growth Objectives would be used to set the performance goals. Baseline data was to be gathered from the previous year's test scores and goals were to be set based on the scores expected during the implementation year. Teachers and administrators were to work together to set the goals. The plan also stated that a goal would be constructed for each of the standards at each of the grade levels. No individual level compensation was given based on student performance.

Parent Satisfaction

Teachers at Canyon Rim were to be awarded pay for community satisfaction based on the results of an online community satisfaction survey developed by Veamos Survey Solutions. The survey was given to parents in March of the implementation year. Teachers with at least 80% of total responses that indicated that the teacher met or exceeded all expectations were rewarded.

Manila

Quality Instruction

Instructional quality at Manila Elementary School was measured with one team-based measure and two individual measures. The team-based measure was evaluated with a 26 item rating scale: the Alpine District Collaboration Incentive Plan. The scale was to be filled out for each team by the team members, an administrator, and a parent committee. The second measure, the Manila Observation Checklist, required 10 5-minute observations of each teacher: five by an administrator and five by another teacher. Feedback on these observations was to be given to the teachers within twenty-four hours of the observation. The final measure of instructional quality was a self-reflection measure. This measure was based on a rubric developed at Manila that included basic level goals up through exceptional level goals. Each teacher was asked to give evidence of the goals that he or she attained by the end of the year.

Student Performance

Student performance at Manila was measured with two team measures and one individual measure. The first team measure was based on CRT scores for students in the second grade or above, a district literacy assessment for the first graders and the USOE assessment for kindergarteners. Progress scores were calculated and grade level teams of teachers were rewarded based on the progress of their students relative to the state expectations. The second team-based measure was similar to the first but used Essential Standards developed at Manila and compensated grade-level teams of teachers for meeting or exceeding school level goals. Individual evaluation measures for each of the teachers were based on

Direct Reading Assessment scores. These scores were compared against the scores from the previous year making student growth the outcome variable on which teachers were evaluated and paid.

Parent Satisfaction

Similar to the other two components, parent satisfaction at Manila was measured at both the team and the individual level. The team level measure was a feedback survey filled out by the parents who attended a presentation on the Manila Professional Learning Community (PLC) process. The individual measure was responses from a Parent Satisfaction Survey. The Manila survey was created from a compilation of parental responses asking what teacher traits and characteristics would satisfy them as parents

Midway

The plan created by Midway Elementary School was strongly aligned with the ideology of the Professional Learning Communities (PLCs). There was a concerted effort on the part of the Midway Elementary planning team to create a plan that was collaborative, in-line with the PLC process and fair. The final plan was embraced by the faculty as being collaborative (96.8% of teachers agreed), supportive of the PLC process (100% of the teachers agreed) and fair (96.8% of the teachers agreed).

Quality Instruction

There were two categories of compensation for teacher quality: PLC requirements and mentorship. To evaluate adherence to the PLC model, teachers were asked to complete six requirements: completing and turning in the essential skills rubric, creating and turning in formative assessments, showing evidence of intervention, showing evidence of enrichment, submitting agendas from the PLC meetings, and completing the quarterly PLC evaluations. Teachers who completed all six requirements received $500, and those that completed five requirements or four requirements received $400 or $200 respectively. To receive compensation for the mentorship component, teachers were asked to work with other teachers as mentors. Teachers who mentored others put together a portfolio of their mentor-related activities and submitted their work to the principal as one of three mentor levels (master—school-level, mentor—team level, or mentor—classroom level) and received additional $100, $200, OR $300 dollars respectively. The principal ultimately decided if the teacher met the criteria for the level of mentorship for which they applied.

Student Achievement

Midway focused their student achievement component on mathematics benchmark scores. Teachers were not rewarded as individuals but rather as a part of a school-wide team or a grade level team. The school-wide team received bonuses if and only if the entire school made Adequate Yearly Progress (AYP) as defined by the No Child Left Behind (NCLB) guidelines. Grade-level teams were evaluated based on the change in percentile on District created benchmark tests in mathematics. Teachers came up with a rubric for measuring progress on these benchmark tests. Tests were given three times during the school year to allow for them to be used as both formative and summative measures.

Parent Satisfaction

Parent satisfaction was measured through a survey sent home with the students. Awards were made based on the percentage of parents who responded to the survey and the average score of each teacher on that survey. There were two other measures of parent or community satisfaction, including the completion of a parent-communication log and participation in PTA sponsored events.

Wasatch

Performance based compensation was not new to Wasatch Peak teachers, as the school had previously awarded teacher bonuses twice yearly. These bonuses were awarded based on a rubric that includes school and individual goals and The Utah Teacher Standards. Based on the contents of their charter, teachers at Wasatch Peak have also received funding for 10 days of professional development outside of the regular school schedule.

Quality Instruction

The quality of instruction requirement at Wasatch Peak had three components: a literacy evaluation, a teacher bonus rubric, and a teacher self-evaluation. Each assessment was to be completed twice per school year. The Literacy Evaluation was a measure that used *The Checklist of Essential Components and Conditions for Effective Reading Instruction* as the evaluation tool. This checklist, an observational assessment developed by Lynn Greenwood of the USOE, was to be completed by an administrator. The Teacher Bonus Rubrics were also completed by an administrator; this measure was a combination of The Utah Teacher Standards and the two charter goals: service learning and Spanish instruction. The final measure of quality of instruction was a teacher self-evaluation based on the Utah Professional Teacher Standards assessment and the charter school goals.

Student Performance

Student performance was assessed using the DIBELS assessment in grades K-2 and Language Arts (LA) CRT scores in grades 3-6. For the purposes of participation in this grant, bonuses were allocated based on meeting the school-wide goal of AYP (20%), grade level goals that were determined by each grade (40%), and individual goals that involved completion of two case studies and meeting at least two of three personal goals (40%).

Parent Satisfaction

A subsection of a larger parental satisfaction questionnaire administered at the school was used to assess parent satisfaction. Teachers received bonuses based on the average endorsement of questions on satisfaction.

EVALUATION DESIGN AND METHODS

As noted in the previous section that outlined the performance-based compensation pilot programs, each of the five sites utilized different measures across the required elements (i.e., instructional quality, student achievement, and parent/community satisfaction. Insufficient common indicators were included across the school plans because each school used different measures. Consequently, we were not able to evaluate the effectiveness of the school performance pay pilot plans per se. Moreover, we could not extrapolate a cross-case comparative analysis of the plans for this reason. Instead, we utilized data that could be compared across sites. In doing so, this evaluation employed a mixed-method approach that addressed the process of the Performance-based Compensation Pilot study over the course of the program. It also addressed the impact that the program had on both teachers and students. Process was primarily assessed through focus groups and individual interviews with administrators and teachers and open-ended responses to teacher surveys. Impact was primarily assessed through a student outcome analysis, a survey given to teachers and a questionnaire filled out by parents. The student outcome analysis the CRT test results as the outcome variable. *It is imperative to note that the individual school site plans did not necessarily include analysis of CRT end-of-year performance data for determination of the performance pay bonuses.*

Student Outcomes

One of the ultimate aims in a performance-based compensation program is to increase student performance and achievement. There is limited research on the efficacy of performance-based pay programs to increase student performance, particularly as measured by criterion reference tests. In fact, previous studies have mostly found weak to no effects, especially for short-term programs lasting less than three year; some have found negative effects. For example, the research by Springer et al. (2011) showed few short term and no long-term gains related to pay for performance in Tennessee. Kingdon and Teal (2010) also showed no relationship between pay for teachers and test scores in India while Eberts, Hollenbeck, and Stone (2002) observed negative academic outcomes in Michigan. The purpose of this section of the evaluation is to explain what, if any, student outcome measures were related to the Performance-based Compensation Pilot as it was conducted in Utah.

Sample

To explore for possible effects of this program, several different analyses were undertaken using different samples of students. Because CRT results are the only available data that are consistent across schools (Ashman students were not included at all in these analyses because CRT testing begins in the 3rd grade and Ashman is a K-2 school and only one cohort from Midway could be used because Midway is a K-4 school) we used a sample of student who had taken CRT tests to measure academic outcome. All students sampled for the student outcome analyses were in the 3rd, 4th, 5th, or 6th grade during the 2011 school year. These grades were chosen because CRT tests were given to students in those grades. The students from performance-based compensation schools included 1,488 students who had CRT scores from the same performance-based compensation school during both the 2010 and the 2011 school years.

The second analysis compared students from the performance-based compensation schools with a matched sample of 1,524 students from five non-performance-based compensation schools. Matched schools were selected for comparison because they were the closest matches to each of the performance-based compensation schools on the variables of district and/or neighborhood (within the same district and neighborhood for public schools and the same neighborhood for charter schools), proportion of minority students, and CRT scores, in that order. Districts or /neighborhoods were the primary indicators for selection because the variance that is accounted for at the district level (about 4%) cannot be statistically accounted for as it is not captured in other predictor variables. Proportion of minority students was used because it is a fixed factor that has a known influence on student growth. CRT scores were selected, even though they are used as a covariate in the analysis, because CRT is the outcome variable.

Because no significant year-to-year differences were found within the students in performance-based compensation schools (analysis 1) and no differences in growth could be detected when comparing performance-based compensation students with students in matched schools (analysis 2), a third analysis was run. The final analysis compared the students from the performance-based compensation schools with the population of all elementary school students from Utah who had CRT scores, from the same school, during the 2010 and 2011 school years. Demographic information about all students used in the three different analyses can be found in Table 9.

Table 9. Demographics Associated With the Sample of Pilot, Matched, and Other Elementary Schools

School Type	Sample Size	Low Income	Low English Proficiency	Special Education	Minority
Performance-based compensation school	1,488	21%	3%	12%	13%
Matched schools	1,524	25%	6%	12%	14%
Population of non-performance-based compensation students	227,790	41%	9%	13%	22%

Measures

Scores from math and language arts CRTs, at that time given to all 3rd through 6th grade elementary school students in the state of Utah, were used in these analyses. The data for these analyses is the USOE Student Test database, which is available to the UEPC as a founding partner with the USOE and other Utah state agencies in the Utah Data Alliance.

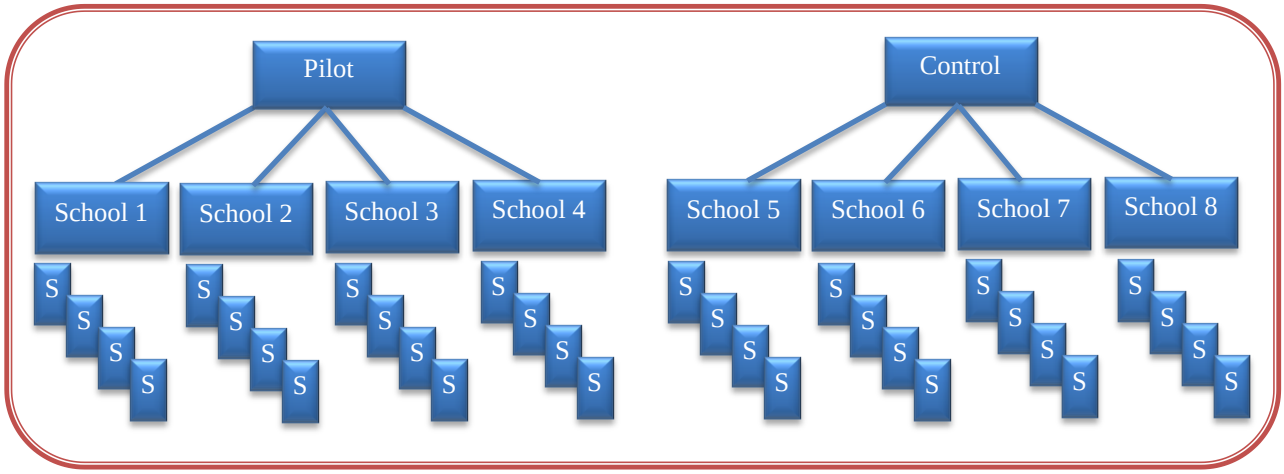
Designs

The first analysis used a within schools design that compared each student's CRT scores in 2010 and in 2011. A split plot design was used to detect differences in performance between 2010 and 2011 and to detect differences in growth between the schools, particularly schools that targeted either language arts or math in their plans.

The second design compared growth in the pilot schools with growth in the matched schools. This design involved a multi-level model with students nested in schools that were nested in treatment conditions. We used a random effects ANOVA that treated students and schools as random variables and the condition (pilot or control) as a fixed variable. The design is represented in

Figure 5 and was used to identify significant differences in academic achievement for the students who were in performance-based compensation schools. The outcome variable at the student level was the difference in CRT scores between the years of 2010 and 2011.

Figure 5. Design for the Student Outcome Comparison



The final analysis was conducted post-hoc to determine if the students in the performance-based compensation schools differed from students from all other elementary schools in the state. The reason for this additional inquiry was the failure to find significant differences between the pilot and control schools in analysis two. This model was run in effort to find any evidence that the Performance-based Compensation Pilot had an effect on student test scores. This analysis used a hierarchical design with all student level covariates on the first level (gender, race, parental SES, LEP and special education designations, and the 2010 CRT scores) predicting CRT scores in 2011. On the second level, the school averages on 2011 CRT scores were considered after controlling all of the first level factors.

The model design specification for analysis 3 was:

- *Level-1 Model*
  2011 CRT score = B0 + B1 (2010 CRT score) + B2*(FEMALE) + B3*(LOWINCOM) + B4*(LIMETED ENGLISH) + B5*(SPECIAL EDUCATION) + B6*(WHITE) + B7* (GRADELEV)+ R*

- *Level-2 Model*
  B0 = G00 + G01(* Performance-based Compensation*) + U0*
  B1 = G10
  B2 = G20
  B3 = G30
  B4 = G40
  B5 = G50
  B6 = G60
  B7 = G70

Teacher Survey

There are well researched issues surrounding performance pay that drove the development of the survey questions used in this evaluation. The issues we used for the survey are supported by research on performance pay and/or strategic pay plans (e.g., Ritter & Jenson, 2010). That is, research supports the following guidelines for the development and utilization of performance pay programs:

1. Programs should generate teacher, staff, and administrator support,
2. Bonuses should be attainable, transparent, substantial, and sustainable,
3. Performance pay programs should foster collaboration,
4. The school structure should support the pay program,
5. There should be multiple measures of teacher effectiveness, and
6. The measures should be perceived as fair.

Besides linking survey questions to current research (Ritter & Jenson, 2010), we focused specifically on the aspects of impact (both on students and on teachers) and the idea of fairness. Research conducted by Mahony, Menter, and Hextall (2004) on an extensive "performance management" program in England (analogous to a pay for performance program in the US) demonstrated that the performance pay program had a negative effect on the teachers emotionally, including "strong feelings of antagonism" and "vitriolic" reactions from the teachers. Besides negatively impacting the teachers emotionally, recent studies have noted the increased focus on student progress, as well as the effort spent in creating and collecting evidence of performance, was seen to as negatively impacting the students as teachers had less time for actual teaching. Our intent with the survey and interviews used in this evaluation was to understand the experience of teachers and to monitor these types on unintended consequences in classrooms.

Sample

All teachers who were in the five pilot schools during the implementation year and remained in the same school the following year were invited to participate in the survey. Of these teachers, just over 60 percent completed the survey and all of the completed surveys were included in this analysis. The breakdown of response rate for each school is included in Table 10.

Table 10. Teacher Response Rate by School

School	Number of Teachers Contacted	Number of Teachers Responding	Percent of Teachers Responding
Midway	31	21	68%
Canyon Rim	19	12	63%
Manila	28	17	61%
Wasatch Peak	17	10	59%
Ashman	24	13	54%
Total	**119**	**73**	**61%**

Note. Teachers contacted to participate in the survey had been at the school during the planning year, the implementation year, and are still at the school this year.

Instrument

The survey was delivered to teachers in an online version using SNAP software. The survey questions are available in Appendix A.

Procedure

Teachers' were contacted with a link to the survey via their school email accounts. They were given two reminder emails if they had not responded within 3 and 5 days, respectively. The initial emails and reminder emails included contact information that teachers' could use if they wanted to be removed from the email list. Only one of the 119 teachers opted out using this procedure.

Design

Each of the survey questions was analyzed using the technique most appropriate to that specific question. Descriptive statistics were calculated for the closed-ended questions, which considered issues of attainability, transparency, substantiality, impact and fairness. The analysis of responses to the open ended questions was primarily focused on impact and fairness issues and contributed to the qualitative analysis for this evaluation. There were three additional evaluation questions for which data were analyzed quantitatively. The additional research questions were: 1) Did teachers who modified their instruction as a result of this program think that the program had a positive impact on students? 2) What factors were correlated with teachers' perceptions of fairness? and 3) What factors do teachers think a fair evaluation would consist of?

Parent Questionnaire

The parent satisfaction questionnaire used in this evaluation was created for several purposes. One use of the questionnaire purposes was to pilot questions for future evaluations of performance-based compensation programs. In an article by Wolstetter, Nayfack, and Mora-Flores (2002) they suggested using items from national surveys to facilitate comparisons between different populations. In their survey, they used a question from the Kappan poll that asks parents to grade schools on an A through F scale just like students are graded. We included a "grading schools" question as well as a "grading teachers" question to see how much of the total variance in all school and teacher items could be accounted for with just those questions. We also used the specific school and teacher questions to determine if there was a difference between satisfaction with the school and satisfaction with the teacher. This was done to determine the validity of asking parents about teachers. If a teacher satisfaction survey measures parent attitude in general and satisfaction with the teacher specifically, there is little point in using these kinds of surveys.

Because the work of Friedman, Bobrowski, and Markow (2007) suggests that communication is the most important non-academic factor related to parent satisfaction we asked a number of questions about the amount of communication that parents had with the teachers. The amount of communication is presented in the results section as descriptive baseline statistics and also correlated with teacher satisfaction scores to help us understand the relationship between communication and satisfaction. Teachers were quite concerned that parents who had not been in the classroom were instrumental in evaluating them for pay. To investigate this concern we compared the responses from parents who had and had not been in the classroom.

Sample

All parents with students in the pilot schools this year were invited to take the Parent Survey. This eliminated from the sample the parents who had students in a terminal grade at the performance-based compensation schools if those parents did not have a younger child currently in the school. Approximately 15% of all parents responded to the parent satisfaction questionnaire. The percentages varied widely with a range between 7% of the parents responding and 37% of the parents responding. The numbers of parents from each of the schools who responded to the survey are included in Table 11.

Table 11. Parent Response Rate by School

School	Number of Students in 2010-2011	Number of Parents Responding	Approximate Proportion of Respondents
Ashman	514	35	7%
Canyon Rim	530	109	21%
Manila	820	82	10%
Midway	509	55	11%
Wasatch Peak	375	138	37%
Total	**2748**	**419**	**15%**

Instrument

The questionnaire was delivered to parents online using SNAP software. The questions used are available in Appendix B.

Procedure

Due to familiarity with the parents, schools newsletters, emails, and notes home were used to distribute the link for the parent questionnaire. Different schools contacted parents differently, depending on the primary systems they had in place for communicating with parents. Most schools sent home notes with the students letting parents know about the survey and directing the parents to either a link to the survey on the school website or directly to the survey link if no school website was available. The survey was scheduled to stay open for two weeks but at the end of the two weeks there were too few responses so notes were redeployed and the survey stayed open for an additional week. Instructions on the initial note home and on the survey itself explained that the survey was for parents who had students in the school LAST year (the 2010-11 school year). Parents who did not have a student in the school last year were thanked for their time but invited not to take the survey.

Design

Each of the survey questions was analyzed using the technique most appropriate to that specific question. Descriptive statistics were calculated for each of the close-ended questions; summated rating scales were created from the school and teacher specific questions resulting in a scaled variable appropriate for inferential statistics. Factor analysis was used to determine if school and teacher satisfaction represented different factors. Non-parametric methods were used to evaluate the rankings given to the different possible evaluation criteria and to compare parent and teacher ranking.

RESULTS

Table 12 provides a brief overview of results for all outcome variables for each participating school. The results of each of the outcomes are described further detail in the body of this results section, including a discussion of the student achievement outcomes, teacher survey results, parent questionnaire results, and a summary of the interview and focus group data.

Table 12. Summary Statistics for all Outcome Variables by School

School Name	Payout Statistics	Average Teacher Satisfaction Score	Average Parent Satisfaction	Average CRT Scores 2010 - 2011
Ashman	School Median: $1582	3.31	Satisfied with Teacher: 4.69	Language: N/A
	School Average: $1561		Satisfied with School: 4.29	Math: N/A
Canyon Rim	School Median: $2000	3.45	Satisfied with Teacher: 4.51	Language: 169.76 - 169.31
	School Average: $1868		Satisfied with School: 4.61	Math: 168.24 - 170.64
Manila	School Median: $2238	3.97	Satisfied with Teacher: 4.37	Language: 168.64 - 168.72
	School Average: $1928		Satisfied with School: 4.38	Math: 169.41 - 169.77
Midway	School Median: $1800	3.75	Satisfied with Teacher: 4.4	Language: 168.96 - 167.97
	School Average: $1716		Satisfied with School: 4.77	Math: 169.64 - 169.26
Wasatch Peak	School Median: $1932	3.82	Satisfied with Teacher: 4.75	Language: 169.06 - 169.93
	School Average: $1889		Satisfied with School: 4.77	Math: 168.85 - 169.83

Compensation Results

Table 13 summarizes the amount of bonuses paid to participating teachers by school, as reported by each school. The average bonus ($1,786) was equivalent to what the average teacher would make in about two weeks of conventional work. Figure 3 presents the methodology by which this estimate was derived.

Table 13. Summary of Bonuses Paid

		Size of Bonus					
	Number of Recipients	**Min**	**Max**	**Median**	**Range**	**Mean**	**Std. Dev.**
District Schools							
Ashman	23	$1,345	$2,057	$1,582	$712	$1,561	$129
Manila	29	$517	$2,617	$2,238	$2,100	$1,928	$690
Midway	31	$1,020	$2,000	$1,800	$980	$1,716	$242
Charter Schools							
Canyon Rim	17	$1,000	$2,500	$2,000	$1,500	$1,868	$556
Wasatch Peak	18	$949	$2,491	$1,932	$1,542	$1,889	$399
Overall	**118**	**$517**	**$2,617**	**$1,800**	**$2,100**	**$1,786**	**$465**

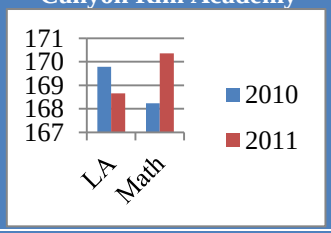
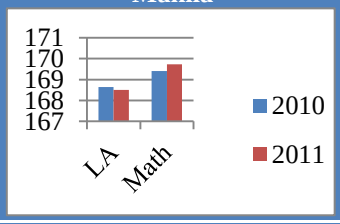
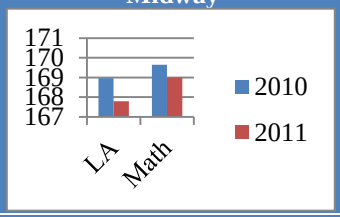
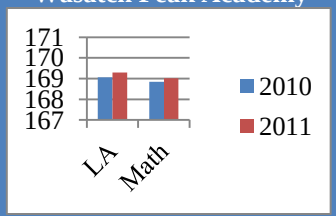
Note. Recipients include all non-administrative faculty listed on reports submitted to the USOE by the pilot schools.

Student Outcome Results

Analysis 1

The first analysis looked exclusively at the students who attended Performance-based Compensation Pilot schools during both the planning and the implementation years. Table 10 presented earlier describes the sample. Table 14 shows the average language arts and math CRT scores for the sample of students from the performance-based compensation schools (Ashman, Canyon Rim, Manila, Midway, and Wasatch Peak).

Table 14. Average CRT Scores by School and Grade

School	Grade in 2011	Language CRT Score 2010	Language CRT Score 2011	Math CRT Score 2010	Math CRT Score 2011
Canyon Rim Academy 	Grade 3	170.13	167.20	167.42	169.16
	Grade 4	171.37	168.52	166.31	169.71
	Grade 5	167.89	168.60	169.24	170.11
	Grade 6	169.86	170.32	169.60	172.47
	School Average	**169.79**	**168.66**	**168.24**	**170.36**
Manila 	Grade 3	171.07	169.11	172.83	170.74
	Grade 4	167.88	167.61	167.78	167.79
	Grade 5	167.75	166.72	166.24	169.95
	Grade 6	169.83	170.73	173.10	170.46
	School Average	**168.64**	**168.51**	**169.41**	**169.73**
Midway 	Grade 3	168.40	166.54	172.11	168.07
	Grade 4	167.71	169.14	165.64	170.01
	School Average	**168.96**	**167.80**	**169.64**	**169.00**
Wasatch Peak Academy 	Grade 3	173.38	171.10	178.21	167.50
	Grade 4	168.00	169.22	165.21	169.41
	Grade 5	167.24	168.14	167.86	171.14
	Grade 6	167.65	169.17	166.94	167.13
	School Average	**169.06**	**169.30**	**168.85**	**169.02**

Note. The **school average** statistics include the average language or math scores for **all** students taking the CRT test in that school year. These averages include results from different student from year to year as the 1st graders in 2010 are not tested in 2010 but are included in the 2011 results as 2nd graders and the 2010 6th graders (and Midway 4th graders) are not included in the 2011 results because they have graduated from the school. Grade level statistics compare the same students from year to year. Ashman is not included in this table or these analyses because CRT testing starts in the 2nd grade and Ashman is a K-2 school.

These descriptive results, as well as the results from the paired samples t tests, did not show a consistent pattern of change from 2010 to 2011 during the implementation of this program. For example, Manila elementary school saw math scores go down significantly for the 3rd graders, stay the same for the fourth

graders, go up significantly for the 5th graders, and go down significantly for the 6th graders. The only statistically significant results at the school level showed the language arts scores at Midway went down slightly but significantly and the math scores at Canyon Rim went up significantly.

The split-plot analysis showed no significant difference between either language arts or math scores from 2010 to 2011. There was also no overall difference between scores at each of the schools. There was a significant difference between the schools when it came to the amount of change that each school recorded between 2010 and 2011.

Analysis 2

The second analysis considered the change from 2010 to 2011 experienced by students in the five performance-based compensation schools compared to the change experienced by students in the five schools that were matched on demographic factors of district or neighborhood, percentage of students of color, and average CRT scores. Table 9 from the methods section presented the Performance-based Compensation and Matched schools. Inferential statistics were calculated and each of the pilot schools was compared with its matched school. The analysis yielded results for three different effects. Two omnibus, or over all effects, (one that tested for overall differences between 2010 and 2011 scores and one that tested for overall differences in year-to-year change between all of the schools) and one interaction effect that tested for differences in growth (change from 2010 to 2011) between students from pilot schools and the students from matched schools. Although the overall effects are of some interests (i.e., do students, in general, change significantly over time?) it is the interaction effect that is the main focus of this analysis since this is where the influence of the Performance-based Compensation Program would manifest.

The results of the model that used language arts scores as the outcome variable showed that overall there was no significant difference between scores in 2010 and in 2011. The overall school effect was also significant, meaning that there was significant year-to-year change in CRT results for the different schools. The interaction effect, however, was not significant indicating that the students from pilot schools and the students from the matched comparison schools experienced the same amount of growth between 2010 and 2011.

The results of the model that used math scores as the outcome variable showed that overall there was a significant difference between scores in 2010 and 2011, with scores in 2011 being about 1 point higher than scores in 2010. There were also significant between school differences between the 2010 and 2011 scores meaning that some schools changed more than others. As with language arts, the most important research result from this analysis was the interaction effect because it tested for differences in growth between pilot and matched comparison schools. Again, there was no significant difference in math scores that could be attributed to the students attending either a pilot or a comparison school. Results for all research questions answered through analysis 2 are presented in

.

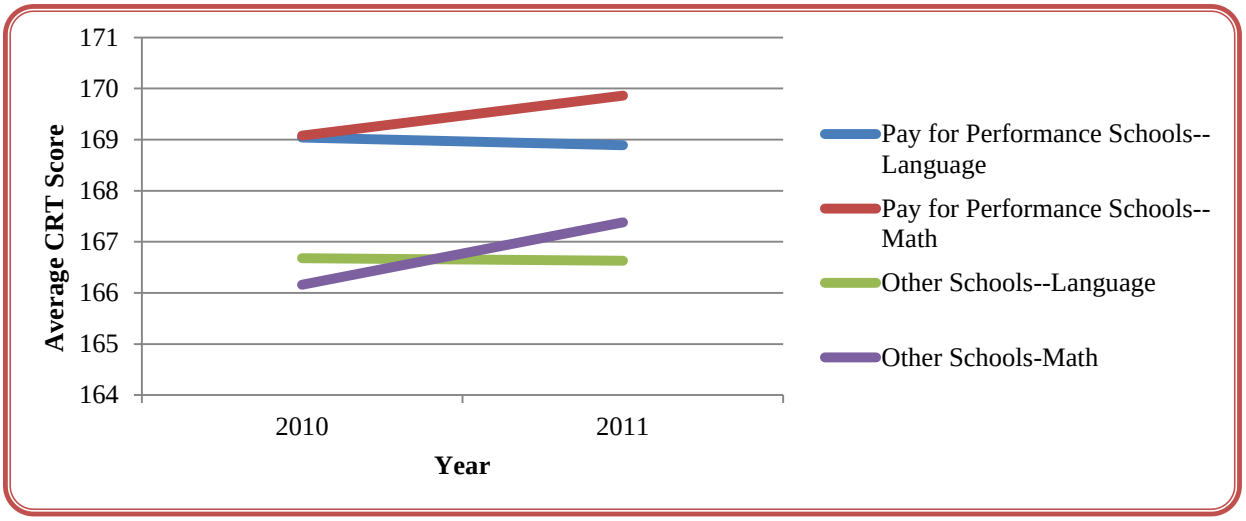
Table 15. Research Questions and Answers from Analysis 2

Language Arts	
Are scores different from 2010 to 2011?	NO
Do scores change differently depending on the school?	YES
Is growth different for the pilot schools than for the matched comparison schools?	NO
Math	
Are scores different from 2010 to 2011?	YES (2011 scores higher)
Do scores change differently depending on the school?	YES
Is growth different for the pilot schools than for the matched comparison schools?	NO

Analysis 3

The third analysis was run to compare the students in the Performance-based Compensation Pilot schools with all other students in the state, accounting for all demographic variables and baseline CRT scores. This was not the primary analysis because the students from the pilot schools differed from the rest of the population on every variable (i.e., ELL, socio-economic status, race, test scores, etc.) with the exception of proportion of females and proportion of students in special education. Figure 6 shows the differences in the math and language arts scores in 2010 and in 2011. In this figure, the bars for the pilot school (red bar for math and blue bar for language arts) are both higher than the bars for the other schools in the state. What is of interest in this analysis, however, is the slope of the bars. The red bar and the purple bar both have equivalent slopes as do the blue and the green bars.

Figure 6. Comparison of CRT Schools for Sample and Population



When the model was run, there were no significant differences in growth for either language arts scores or math scores between students in schools. Complete model specification and the results tables for this analysis are available in Appendix C.

We evaluated evidence of an effect within students from pilot schools, between pilot school students and matched school students, and when pilot students were compared to the population. Given our analyses, there is no evidence to support the hypothesis that students in schools that had performance-based compensation pilot programs experienced any change in growth on the annual CRT scores as a result of these programs.

Teacher Survey Results

Frequency statistics were calculated for the fixed response items and are presented here as the percent of teachers who agreed with each statement. Agreement for the first five questions was indicated by answering "Yes." Agreement for the remaining statements was indicated by teachers either responding with "agree" or "strongly agree" on a Likert scale with five rating points. Percent of teachers agreeing at each school and overall are included in Table 16.

Table 16. Percent of Teachers from Each School Who Agreed with Each of the Forced Response Questions on the Survey

	Ashman	Canyon Rim	Manila	Midway	Wasatch Peak	Overall
Question	***Proportion of teachers agreeing with statement* (Total number of respondents who agreed)**					
Planning						
Did you help to develop the performance-based compensation program at your school?	69%	67%	83%	86%	90%	**80%**
Modification of instruction						
Did you modify your teaching of mathematics in any way as a result of this program?	25%	54%	40%	81%	44%	**51% (32)**
Did you modify your teaching of Language Arts in any way as a result of this program?	42%	0	53%	28%	40%	**38% (22)**
Modification of interactions						
Did you modify your interactions with the students in any way as a result of this program?	15%	50%	41%	29%	20%	**32% (23)**
Did you modify your interactions with parents in any way as a result of this program?	39%	58%	53%	62%	60%	**55% (40)**
Transparency						
I understood how to receive a reward through my school's incentive program	69%	92%	82%	86%	90%	**84% (61)**
Fairness						
I felt like the distribution of performance-based compensation funds was fair.	62%	42%	65%	48%	70%	**56% (41)**
Financial Reward						
I expected to be rewarded through this program	54%	83%	82%	86%	80%	**78% (57)**
I was financially rewarded through this program	92%	92%	94%	86%	90%	**90% (66)**
The rewards offered were of sufficient value	31%	36%	76%	57%	40%	**51% (37)**
Overall Impact						
This program improved the learning experience for students	31%	50%	65%	67%	30%	**52% (38)**

	Ashman	Canyon Rim	Manila	Midway	Wasatch Peak	Overall
Question	***Proportion of teachers agreeing with statement* (Total number of respondents who agreed)**					
This program improved the school experience for teachers.	31%	25%	77%	57%	30%	**48% (35)**

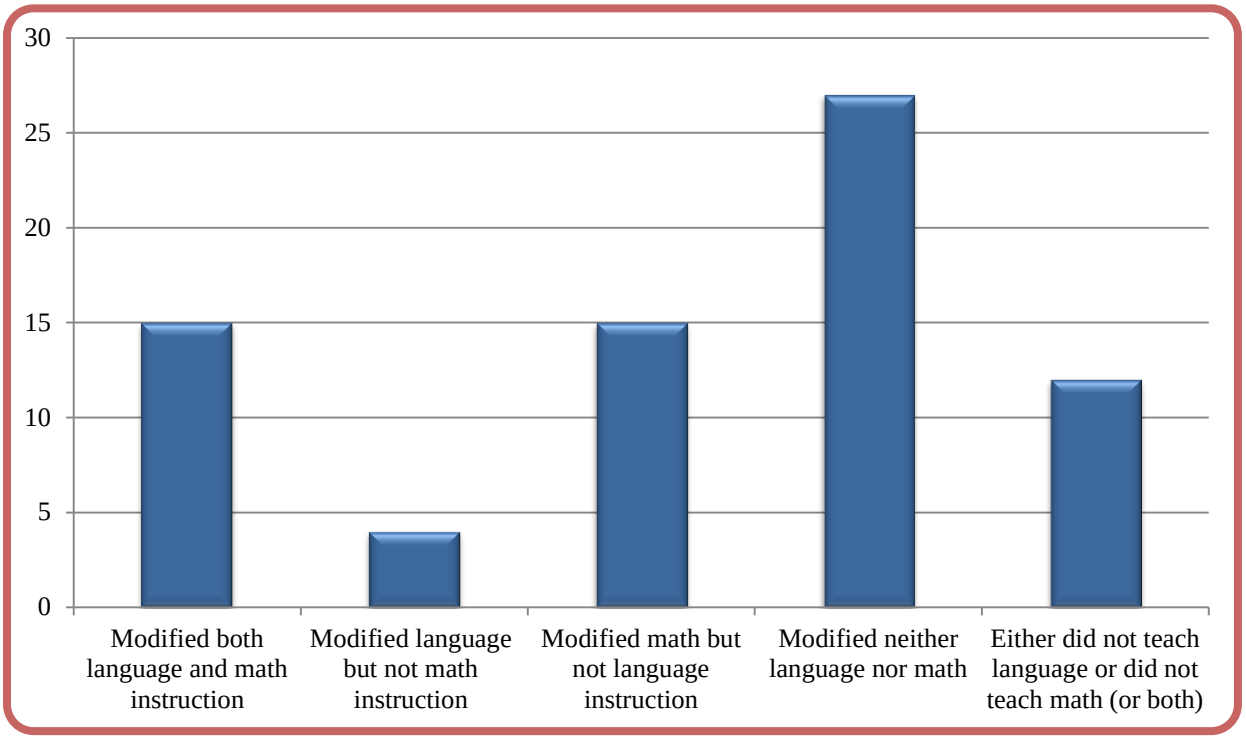
Note. Agreement was determined by answering "Yes" to the yes or no questions or answering either "agree" or "strongly agree" with Likert scored items

Results indicate that teachers, for the most part, helped to develop the performance pay program in their schools, understood the program, expected to be rewarded, and in the end were rewarded through the program. Teachers were divided when it came to whether or not the bonuses were substantial with 51% indicating that they were sufficient, 29% giving a neutral response, and 20% indicating that the bonuses were not sufficient. Teachers were similarly divided on the fairness question with 56% agreeing that the program was fair, 27% responding neutrally, and 16% disagreeing. Overall, this descriptive analysis shows that the money offered during the performance-based compensation pilot program probably would not be enough over a longer period. Results also indicate that there is room for growth when it comes to the perceived fairness of these programs.

Impact

The lowest agreement frequencies were related to the impact of the program. As indicated in Figure 9, about half of the teachers (51%) agreed that they changed math instruction as a part of this program and significantly less than half (38%) agreed that they changed language arts instruction. When these two types of instruction were considered together we found that of the 61 teachers who taught both math and language arts about 25% of them modified instruction for both topics and 44% did not modify instruction for either discipline.

Figure 9. Frequencies of teachers who modified instruction of math and/or language (or not modifying either)



Teacher opinion was evenly split when it came to the impact that the program had on student learning and on the experiences of teachers in the program with just more than half of the teachers (52%) agreeing that the program improved student learning and just less than have agreeing that it improved the school experience for teachers (48%). We were particularly interested in the way that teachers who had modified their instruction responded to this question. Of the teachers who said they had modified their teaching, 74% agreed that the program had a positive impact on student learning. Of the teachers who did not, only 26% agreed that the program had a positive impact on student learning. Similarly, 61.5% of the teachers who modified teaching agreed that the program was beneficial to teachers compared to 32% agreement for teachers who did no modification.

A non-parametric statistical test (the Mann-Whitney) was run to test for significant differences between the agreement ratings from the teachers who had modified instruction and the teachers who had not. Not surprisingly, the results were significant with the 39 teachers who did alter their instruction being significantly more likely to agree that the program improved learning for students ($p=.001$) and significantly more likely to agree that the program improved the school experience for teachers ($p=.03$).

In what ways did teachers modify their language and math instruction?

Teachers who said that they had modified their teaching in either math or language were asked for specific examples of how they modified their teaching. The responses were categorized and the most frequent responses were related to increased focus on data, evaluation, and or assessments and a greater focus on goals and objectives was listed second most often. Typical responses, in order of frequency included:

- Use of data, evaluations or assessments
- Increase focus and use of goals and/or objectives

- Gave students more practice
- Used more differentiated instruction
- Made teaching or instructions more explicit
- Used classroom aides differently
- Spent more time planning
- Used curriculum maps
- Integrated disciplines
- Had different priorities
- Worked as a team

In what ways did teachers modify their interactions with students?

The lowest agreement rating for any item on the survey was in response to the statement about modifying interactions with students. Only 32% of the teachers said that they had modified their interactions with their students. Teachers who did agree with the statement that they had modified interactions with students gave examples that generally fell into one of three categories: (1) more one-on-one time, (2) more positive interactions (e.g., more patient, more positive, using more praise, etc.), and (3) using data more to give students feedback. These categories were mentioned five, five, and four times, respectively.

In what ways did teachers modify their interactions with parents?

Interestingly, most teachers (55%) agreed that they had changed their interactions with parents. Virtually every one of the examples given were related to increased or improved communication with the majority of teachers describing and increase in the data they shared with the parents and the specific feedback they were able to give. Other typical responses to the open-ended question about how interactions with the parents were modified were:

- Increased frequency of communication
- Surveys
- Development or use of a class website
- More notes sent home
- Kept communication records/increased awareness
- Made or returned phone calls more diligently
- Class newsletter
- Collaborative presentations to parents

What were the positive aspects of this program?

In reviewing the responses to the question, "what were the positive aspects of this program?" It was apparent that increased collaboration, monetary reward, and improved focus and/or goal setting were the most salient positive aspects of this process. Other responses, in order of frequency, are presented below:

- Collaboration
- Receiving the reward
- Setting and tracking goals
- Increased focus, awareness, and/or motivation
- Increased communication with or feedback from parents
- Making observations or being observed
- Better access or use of data
- There were no positive aspects
- Encouraged better teaching
- The program was motivating

Single responses to this question included the rubrics, accountability, that it was teacher designed, increased school pride, and that the premise was good.

What were the greatest drawbacks of the program?

Responses to the question, "what were the drawbacks of this program?" were either related to the negative impact on the teachers or perceived unfairness' associated with the program. Table 17 shows a breakdown of responses to this question.

Table 17. Responses to the "Greatest Drawbacks" Question

Category	Response related to impact or fairness	Frequency
Extra work	Impact	16
Non-collaborative/competitive	Impact	10
Unequal distribution for specialists	Fairness	7
Focus on test scores (teaching to test)	Impact	6
Progress determined by non-teacher factors	Fairness	6
Reliance on Parent surveys	Fairness	5
Negative emotional impact on teachers	Impact	4
Not all teachers could receive highest reward	Fairness	2
Total number of teachers responding to this question		**65**

Note. There were a number of single response categories relating to duration, amount, scope, parent surveys, new teachers, distribution, and subjectivity.

Fairness

We created a simple correlation matrix to determine which factors were correlated with teachers thinking the program was fair. The results showed six factors to have significant positive relationships with the fairness question and two factors to have small to medium negative relationships (see Table 18). The strongest correlation was with the statement, "I was financially rewarded through this program." A cross-tabulation showed that there were only eight teachers who indicated that they were not rewarded through this program and, of those eight, six of them indicated that the distribution of funds was extremely unfair. The strength of the relationship between these two factors is almost entirely driven by this small group of teachers. The response matrixes for the other correlates were much more balanced and can be interpreted more directly.

Table 18. Survey Items that Significantly Correlated with the "Fairness" Question

Response Correlated with Fairness Question	Pearson's Correlation Coefficient	Interpretation of Correlation Coefficient.
I was financially rewarded through this program.	.673	Strong positive relationship between being rewarded and the perception of fairness. Skewed by an outlying group.
This program improved the school experience for teachers.	.637	Strong positive relationship between believing the program improved the school experience for teachers and the perception of fairness.
I understood how to receive a reward through my school's incentive program.	.619	Strong positive relationship between understanding how to be rewarded through the program and the perception of fairness.
This program improved the learning experience for students.	.592	Strong positive relationship between believing that the program was good for students and the perception of fairness.
I expected to be rewarded through this program.	.527	Strong positive relationship between the expectation of reward and the perception of fairness.
The rewards offered were of sufficient value.	.494	Strong positive relationship between thinking the reward was sufficient and the perception of fairness.
Did you modify your interactions with parents in any way?	-.256	Medium negative relationship between teachers who modified their interactions with parents and thinking the program was fair. (Teachers who modified behavior were less likely to view the program as fair.)
Did you modify your interactions with the students in any way?	-.236	Small negative relationship between teachers who modified their interactions with the students and thinking the program was fair. (Teachers who reported a modification of behavior were slightly less likely to say the program was fair.)

The positive relationships seen in Table 18 are intuitive as they show teachers who thought the program was fair also thought the program was effective, the rewards were substantial, etc. The negative correlations are less intuitive. These relationships show that teachers who thought the program was fair were less likely to modify their interactions with students or teachers and conversely teachers who did modify their interactions were less likely to see the program as fair. In the qualitative analysis it was apparent that a number of the teachers saw the requirements of the program as taking time away from their teaching responsibilities. It is possible that these negative correlations are manifestations of that same sentiment.

As another measure of fairness, teachers were asked to rank seven different indicators from 1 (highest rank) to 7 (lowest rank) in terms of their importance for determining pay for teachers. The results of this ranking, shown in Table 19 showed a strong preference for pay based on instructional quality and/or the academic performance of their students and a strong preference to *not* be rewarded based on test scores or on parent satisfaction. The factors of classroom management, teachers' experience, and teachers' education, all in the middle of the rankings, did not show sufficient differentiation to say that one was favored over the other. It is interesting that 12 of the 73 teachers (16%) ranked instructional quality as the least important factor, given its ranking on every other factor (mode, median, mean and number of

"most important" rankings) as the most important factor. This certainly shows that the opinions of teachers very widely when it comes to how teacher pay should be determined.

Table 19. Rankings by teachers of Factor Importance for Determining Teacher Pay

	Rank	Median	Mean	Number of "Most Important" Rankings	Number of "Least Important" Rankings
Instructional quality	1	2	2.88	30	12
Student academic performance	2	3	3.7	16	8
Classroom management	3	4	3.78	4	4
Teacher's experience	3	4	3.95	4	4
Teacher's education	3	4	4.01	10	8
Parent satisfaction	4	5	4.81	6	18
Test scores	4	5	4.88	3	19

Note. The better the ranking, the lower the scores

When asked what other factors should be included in performance pay the most popular responses were to pay for collaboration or team work and to pay based on the individual differences between students (for teaching ELL, special education, low performing, or behavior students). Other ideas beyond the to-be-ranked factors were overtime pay and pay based on student evaluations.

As previously mentioned, fairness issues also came up when asked about the drawbacks of the program (See Table 17). The most common fairness issues were the non-inclusion of specialists or non-core teachers (mostly these responses came from a single school), student factors beyond the teacher's control, the use of parent surveys, and some pay issues (also from a single school where not everyone could qualify for one of the requirements.

General Impressions

One item on the survey was an open opportunity for the teachers to share anything they wanted concerning the close-ended questions. Thirty-three out of the 73 teachers responded to this open-ended invitation. The responses to this item were among the most passionate of all the responses that teachers gave. Responses were coded as generally positive, generally negative, or neutral. Unfortunately, there were significantly more negative (14) than positive (8) comments. The negative comments were sub-coded into those related to the impact of the program (8), the fairness of the program (4), or generally negative responses (2). Examples from each of the coded categories are included in Table 20.

Table 20 Examples of positive, negative, and neutral statement made when teachers were asked for general comments.

Code	Comment
Positive	*I liked having people observe me for a few minutes at a time because it gave me feedback to reflect upon.*
	The process of developing and implementing the plan was cohesive and collaborative.
Negative	*The performance pay is the most ridiculous thing that I have ever been a part of. It is truly not worth it.*
	I haven't been a fan of the performance pay program.
Negative--fairness	*I don't like a pay for performance concept; I get a majority of the difficult students each year; why don't I get extra compensation for that?* *How can I be held responsible for a child that does not come to school on a regular basis?*
	…on my team at least only a couple of us worked on fulfilling the assignments, but everyone was rewarded.
Negative--impact	*The pay for performance study caused divisiveness and stress in our school.*
	This was a very stressful and not very productive use of teacher time.
Neutral	*I try to do what's good for students to succeed. The financial reward wasn't my incentive.*
	I am sure this program would be more effective in a school where they need more incentives.

Unlike the other questions that asked for specific feedback, this item was left open for teachers to share anything they wanted or not at all. The particular quotes were selected because they were representative of what the teachers in the various coding categories had to say and may be representative of the opinions generated by performance-based compensation programs in general.

Parent Questionnaire Results

Maintaining consistency with the Kappan National Survey, we asked parents to grade both schools and teachers with the same A through F scale used to grade junior high school and high school students. Frequencies of grades given for each school are presented in Table 211 along with a Grade Point Average (GPA) score created by weighting A's as 4.0, B's as 3.0, etc. as is done in most schools. The grades were generally high, with school GPAs ranging from 3.46 to 3.93, and teacher GPAs ranging from 3.65 to 3.89. Overall, there was no significant difference between the grades given to schools and the grades given to teachers. However, there were significant differences in two of the schools with Ashman teachers graded more favorably than their school and the Canyon Rim Academy graded more favorably than its teachers. The differences between schools for both school grades and teacher grades were also significant ($p<001$ and $p=.025$, respectively).

Table 21. Parent Grading of Schools and Teachers.

		School Grades		Teacher Grades	
School	**Grade**	**Frequency**	**Percent**	**Frequency**	**Percent**
Ashman Elementary	A	23	65.71	32	91.43
	B	6	17.14	2	5.71
	C	5	14.29	1	2.86
	D	1	2.86	0	
	F	0		0	
	GPA	**3.46**		**3.89**	
Canyon Rim Academy	A	98	91.59	93	85.32
	B	9	8.41	10	9.17
	C	0		4	3.67
	D	0		2	1.83
	F	0		0	
	GPA	**3.84**		**3.78**	
Manila Elementary	A	57	69.51	63	76.83
	B	18	21.95	12	14.63
	C	7	8.54	5	6.10
	D	0		1	1.22
	F	0		1	1.22
	GPA	**3.61**		**3.65**	
Midway Elementary	A	50	90.91	43	78.18
	B	4	7.27	8	14.55
	C	1	1.82	3	5.45
	D	0		1	1.82
	F	0		0	
	GPA	**3.89**		**3.69**	
Wasatch Peak Academy	A	129	93.48	122	88.41
	B	8	5.80	15	10.87
	C	1	0.72	1	0.72
	D	0		0	
	F	0		0	
	GPA	**3.93**		**3.88**	
Total	**School GPA=3.80**			**Teacher GPA=3.78**	

School and Teacher Satisfaction

To further gauge parent satisfaction with schools and teachers, we presented four statements about the schools and seven statements about the teachers for parents to respond to with responses ranging from "Completely agree" to "Completely disagree." In spite of the small sample sizes from most of the schools, there was an adequate total number of responses (419) to warrant a statistical analysis. The first analysis was a factor analysis, run to determine whether or not the school and teacher questions represented separate factors.

The results of the factor analysis (presented in Table 22) showed two distinct factors: one for the first four statements, which were specific to the schools and one for the remaining seven statements, which were specific to teachers (accounting for about 51% of all variance). There were differences in the overall responses (e.g., some parent may have agreed with one item, while another strongly agreed with it and both may have disagreed with another) these differences in responses are called variance and the two factors (school and teacher) that emerged from this analyses accounted for 31% and 52% of the variance

respectively. Taken together, school and teacher factors accounted for 83% of the variance, leaving vary little variance to be explained by other factors, such as parent mood, regional differences, etc.

The component matrix for this analysis is presented in Table 22 and shows the factor loading for each item, or the correlations between the item and the factor. Looking at the component matrix, the highest factor loading for the schools factor was on the statement "My child's school was a safe place." The statement with the highest factor loading for the teacher factor was, "The teacher motivated my child to learn." This items best represented the respective factors and would be recommended items for future parent satisfaction surveys.

Table 22. Results from the Factor Analysis of Survey Items

	Factor	
	Teacher	**School**
I felt welcome when I entered the school.	0.342	0.818
The principal at my child's school was an effective leader.	0.28	0.797
My child's elementary school was a safe place.	0.289	0.876
My child's school was clean and well maintained.	0.285	0.761
The teacher helped me to understand what my child was learning.	0.812	0.407
The teacher went "above and beyond" in working with my child.	0.883	0.28
The teacher had a positive attitude about the school.	0.851	0.38
The teacher motivated my child to learn.	0.917	0.295
The teacher had high expectations for my child.	0.856	0.322
The teacher used multiple approaches to help my child learn.	0.902	0.27
The teacher cared about my child.	0.871	0.351

To assess these questions for use in future surveys we also used an ordinal regression technique that predicted school or teacher grade from responses to each of the school or teacher statements. There were serious violations of the assumptions of ordinal regression due to the number of empty cells so the results of regression analysis should only be used descriptively. The results showed that the statement "My school's principal is an effective leader" was the best predictor of the overall school grade and the statement "The teacher motivated my child to learn" was the best predictor of the overall teacher grade.

The results from the regression and factor analyses show that there was a difference between the responses of parents to school-based and teacher-based statements. This supports the validity of surveying parents about teacher satisfaction and refutes the idea that teacher satisfaction scores are just a measure of general parent attitude. The overall school and teacher satisfaction scores are highly correlated ($r = .652$; $P<.001$) so it may be prudent to use school satisfaction scores to control for parental attitude when using parent satisfaction scores to determine teacher pay. Good school satisfaction items include statements similar to "my child's principal is an effective leader" and "my child's school is a safe

place" as these items showed good psychometric qualities (i.e., they correlated with the factor and with the school grade, respectively, and they were not too strongly (r = .697) correlated with each other).

These analyses also support the inclusion of a question or statement similar to "The teacher motivated my child to learn" in future surveys of parents. This statement had the strongest correlation with the "teacher factor" and was the strongest predictor of school grade. If future research is done on parent satisfaction surveys, we recommend pilot testing to find more items that are uncorrelated with the current item set but are correlated with the grading teachers' item. The use of an overall grade along with the specific items, as was done in this survey, is a good practice since the total sum of items used in this analysis only accounted for about 40% of the variance in the grade question (and vice-versa). This means that 60% of the variance was unique to the respective items or item types.

Communication

Of the parents who responded to the questionnaire, 99% had some form of quarterly communication with their child's teacher and 68% of parents communicated with the teachers more often (monthly or better). The most frequent form of communication were face-to-face meetings (half of all parents met with teachers monthly or better) and the least frequency form was via phone calls (half of all parents did not have quarterly phone communication. Frequencies of all communication methods are included in Table 23.

Table 23. Frequencies of Communication by Method

Form of Communication	Response	Frequency	Percent
Face-to-face	Less than once per term	9	2.1
	Once per term	181	43.2
	Monthly	107	25.5
	Weekly	112	26.7
	Daily	10	2.4
Written	Less than once per term	98	23.4
	Once per term	125	29.8
	Monthly	100	23.9
	Weekly	89	21.2
	Daily	7	1.7
Phone	Less than once per term	303	72.5
	Once per term	62	14.8
	Monthly	44	10.5
	Weekly	8	1.9
	Daily	1	.2

One of the most serious concerns that teachers have about using parent satisfaction scores to pay teachers is that the teachers are being evaluated by parents who have little, if any, first-hand knowledge about what happens in the classrooms. These concerns are somewhat validated by the results from this questionnaire as about 30% of the parents who responded to this survey had (presumably) only communicated with the teacher at parent teacher conferences.

The other factor that we were interested in investigating with this research was the correlation between communication and parent satisfaction scores. We found a small but significant correlation (r = .198, p<.001) between the amount of communication that parents had with teachers and the teacher satisfaction rating. This relationship was expected given the work of Friedman, Bobrowski, and Markow (2007) and may indicate that what satisfies parents is communication. The relationship between

communication and school satisfaction was also significant (r = .109, p=.027), which may support an alternative explanation as questions such as "how clean was the school," and "how safe was the school" should not be impacted by whether or not the teachers and the parents had adequate communication with the teacher. For example, this could show that parents who are more willing to participate have a higher opinion of the child's education, in general. The higher ratings may not be related at all to the teacher's attempts to communicate but entirely driven by the parent. Further research should be done in this area to understand the relationship between communication and satisfaction, especially if parent satisfaction is to be used as a determinate of teacher pay.

Program Impact

When asked if they were aware that a performance-based compensation program had been implemented at their child's school, less than half (44.5%) of the parents said yes. Knowledge of the program varied significantly between schools with a range of 24.4% to 56.9% of the parents aware of the program. The parents who indicated that they were aware of the program were asked what, if any, impact they thought the program had on their students. Responses to that question were coded as Positive Impact, Not Sure, or No Impact and the frequency percentages for each response, by school, are provided in Table 24. No parents who responded to the questionnaire thought that the program had a negative impact on their students.

Table 24. Responses to Question about Impact

	Did Program Have an Impact?			Number of Responses
School	**Yes**	**No**	**Don't know**	
Ashman	30.77%	38.46%	30.77%	13
Canyon Rim	36.00%	18.00%	46.00%	50
Manila	40.00%	25.00%	35.00%	20
Midway	80.00%	10.00%	10.00%	10
Wasatch Peak	49.21%	6.35%	44.44%	63
Average	**44.23%**	**15.38%**	**40.38%**	**156**

What kind of impact did the Performance-based Compensation Program have on your child?

When asked what impact the program had on their students, the parents who responded positively typically did not address the impact on their student, but the perceived ideal impact on the teacher. The number one response was that the program served to motivate the teachers. Response categories given by parents, in order of frequency, included:

- Motivation
- "Goodness" of rewarding good teachers
- Increased collaboration among teachers
- Increased focus on student needs
- Differentiated instruction
- Increased communication
- Keeps good teachers/outs "bad" teachers

Fairness

Parents were not asked if they thought the program was fair in the same way that teachers were because most of the parents had little, if any, knowledge about the nuances of the program. Parents were asked,

however, to rank the same seven indicators that teachers were asked to rank from 1 to 7 (with one being most important and 7 being least important) in terms of the item importance for determining the pay of teachers. The results of this ranking, which are shown in Table 25, were remarkably similar to the results of the ranking done by the teachers. The only measurable difference between the parent results and the teacher results is that the parents ranked test scores more highly (fourth based on mean rankings) than the teachers did (seventh based on mean rankings). All other rankings were statistically equivalent. Both parents and teachers are in favor of tying performance pay to instructional quality and student performance. Neither parents nor teacher are in favor of tying performance pay to parent satisfaction.

Table 25. Rankings by Teachers of Factor Importance for Determining Teacher Pay

	Rank	Median	Mean	Number of "most important" rankings	Number of "least important" rankings
Instructional quality	1	2	2.77	160	34
Student academic performance	2	2	3.05	123	38
Classroom management	3	4	3.88	27	24
Test scores	4	4	4.34	21	55
Teacher's experience	4	5	4.39	25	57
Teacher's education	4	5	4.47	33	67
Parent satisfaction	5	5	4.99	34	141

Note. The better the ranking the lower the scores

When asked about other factors that should be considered to determine the pay of teachers, the parents responded with a large number of ideas and anecdotes. For instance, some of the different factors that parents suggested should be determinates of pay included: professional development, class size, cost of living, difficulty of subject taught, student characteristics, student progress, individualized attention, student evaluations, principal evaluations, students' attitudes, teachers' attitudes, over time, and the amount of personal money spent on the class.

Focus Groups and Interviews

In addition to the student outcomes, teacher survey, and parent questionnaire results, we also conducted interview and focus groups at the five pilot sites to gather more in-depth, nuanced information about the implementation and impact of the performance-based compensation program. Interviews were conducted with each principal, a representative district administrator or charter support personnel and focus groups were conducted with planning committees, governance committees, parents, and teachers. We begin with a discussion of the overall context that for schools implementing their plan, followed by a discussion of the factors related to planning and implementation. We also discuss perceptions about the influence of the program on instructional quality and parent satisfaction and conclude with a summary of the challenges experienced in this pilot program.

Context

Overall, we found the following with regard to how the school has described the importance of their context for their participation in the program. Specifically, participants reported that:

- Their uniqueness necessitated a tailored plan.
- Their plan provided focus for faculty and the school and oriented schools toward areas of improvement.
- It would be inherently difficult for schools to successfully implement a replication of any of these plans.

Unique Plans Situated in Unique Schools

The five schools selected to plan for and implement the performance-based compensation pilot perceived their environment as a "unique case," particularly as it related to the quality of staff and the ability of the school to development and implement a plan that rewarded teachers. In part, schools attributed their uniqueness to the degree of success that each school had to date in teaching and learning. They also believed this success enhanced their ability to implement successfully the performance-based compensation pilot. Descriptions throughout the interviews reflected this view included remarks by teachers, administrators, and parents that their sites were "*good*" schools with "*strong*" communities or comparisons of their schools to others, such as we are "*a pretty top-notch school in the district.*"

In addition to their uniqueness, participants believed that their engagement with the performance-based compensation program was facilitated by the degree to which it was aligned with other initiatives/efforts. Alignment eased transition to what was a temporarily supported initiative. Moreover, participants concurred that ensuring alignment of the performance-based program to their current efforts helped buffer the experience, as the program "meshed" into existing efforts:

> *The things that we did seemed pretty natural. The organization was tough, yes, and trying to get everything into one portfolio, that was difficult, but they were things that, for the most part, were being implemented anyways. And so it was nice to see we really are working together for the benefit of the students.*

In fact, several noted that it was difficult to disentangle what was attributed to the pilot program and what was attributed to another existing program. For example, a principal noted the mentoring program, which was a similar, complimentary, and pre-existing program:

> *It's hard to separate that from the mentoring because at the same time we were so engaged in the mentoring and we were having people come here and actually that was risk-taking for my teachers. (Principal)*

Parents also agreed that the performance-based compensation pilot program was tightly aligned with what the school was doing anyway.

> *The cause and effect is almost impossible to distinguish because the program was implemented into an existing philosophy at the school that involved high quality performance and success at all levels. And so I don't think you could say necessarily that our success was driven by this, but this was just one more thing to help contribute to the success.*

Difficulty Applying Plans to Other Schools

The schools saw their "uniqueness" as a testament to why considerable thought and deliberation would be needed for other sites to engage in a similar program. For example, the schools further noted the difficulties in implementing a similar initiative in sites that were not as collegial or collaborative. In recognition of the uniqueness of each school, there was overwhelming consensus that individual school plans could not simply be replicated from school to school. As a parent/board member indicated:

> *I don't think you can replicate what's been done here considering the time that was spent and getting the buy-in from all the teachers and everything.*

As indicated earlier, teachers also supported the need for unique programs tailored to particular schools:

> *When we talked about this once we started implementing it and talked about our plan, we decided that this was catered – I mean, it was specifically designed for our school and*

our needs, our community. It would not work at a school that was – did not have the same situations.

Again, the need for adopting unique, tailored programs was reiterated by teachers and others alike who worried that trying to have the same performance-based compensation pilot program at different schools would predispose the program to failure.

I think trying to take this program, our program – and like I said, I mean, they said they wanted to see if educators could do it, and everybody I know, the school's all did it different, but I just – I don't see how – it would have to be fairly similar so that everybody's accountable on the same level, I would think. (Principal)

Focus and Orientation towards Change

Principals reported applying for the pay for pilot program for numerous reasons, including using the program as leverage for change, a mechanism to reward teachers for their instructional quality, and an opportunity to recognize and prompt student growth and student performance.

I want to do these things and I want to push it a little faster. And here's this little high-profile grant thing that I can say, "Oh, you guys we've got to do this, and we've got to this really well because the state's looking at us." And it didn't hurt to have superintendent and people flying in every once in a while to talk to us." Yeah, it did work. It worked to help me move it. (Principal)

Another principal, who discussed the school's efforts to "*create an effective program*," provided insight into the decision to apply to be a performance-based compensation pilot program site. The principal also highlights the difficult task they had:

Wow, this is something I would really like to have my school in. I would like the teachers to be able to have this extra income, extra money, and I could see the potential for it generating a lot of discussion. I was being pretty ideal about it, like this is really going to help our school. All the teachers when we presented it were really excited about it and wrote their letters, and our proposal was accepted. But then, when we started into the work, it was probably one of the hardest things that I've ever done as far as leading our teachers through this because I took the process of doing it really seriously.

In addition to providing leverage for improvement, the performance-based compensation pilot program reportedly also helped with focus. Participants described how the planning and implementation process provided an avenue to discuss qualities of excellent instruction, what counts for student learning, how to assess student learning, and how to set and achieve individual and collective goals. A principal explains the role of the pilot in providing a framework for these efforts:

The framework is helping us define excellent instruction. And a lot of this framework resonates with them already, but because we're not all on the same page with what the terminology means, we're having to learn the framework and we're having to learn the language describing our craft, just like we had to learn the framework and the language of assessment, we're now having to do that in instruction. [Principal]

In addition to an increased "focus," principals also indicated that the pilot program gave them an "organizational framework" by which they could guide changes and focus attention on areas of need. The following is representative of the responses regarding the guidance that the instructional quality, student achievement, and parent satisfaction framework provided.

> *It gave me an organizational framework for my voice with my staff. When you say the 40/40/20, it just fit so nicely to say, "Okay. We've got the test scores, but you guys, we also have this qualitative stuff called our instruction. And we can talk about this and we can make a difference here, and I've got a scale and you can tell me if I'm seeing it when I come into your classroom. [Principal]*

Parents also remarked on the changes that complimented and refined the school's focus:

> *I didn't all of a sudden see all these things start happening or "we've got to do all this stuff to get our performance-based-pay." It was just maybe fine-tuned things that were already in place.*

Planning

An important aspect of this pilot program was the opportunity for participating schools to plan for the implementation of the compensation program. With regard to planning, we found the following key themes:

- Planning committee membership and stakeholder input varied from school to school.
- There was a need for more support during the development of these plans.
- Plans included individual measures to ensure differentiation of pay.
- Plans and planning maintained a focus on collaboration.
- Tiered approaches included both individual and team-based goals.

Planning Committee Participation

Planning committee configurations, including roles and functions of teachers, differed by site as did the indicators selected. While principals recognized that limiting the planning committees could ease the load on teachers, most opted to include staff more broadly in the planning process. The decision to rely on the teachers to create the plans was associated with a number of different benefits and costs. For instance, a principal indicated that the choice to have everyone contribute to the planning served to avoid perceptions that the measures favored those who developed the measures.

> *They did and everybody got part of the pot. Here's what my fear was. If in the first year I selected just a group, they [the staff] would think those were the people that would be at the top of that performance scale. I wasn't willing to do that .....*

In general, teachers' participation in the planning process was viewed as imperative from both the administrators who included them and from teachers who were involved:

> *You created it. You knew. You remembered the blood and sweat and tears that went into arguing over what verbiage to use and what – you know, so it – I don't know. I felt like that – like, when I was handed my sheet saying, "This is the X-amount of dollars you'll be rewarded," I didn't have any questions as to why or what or how come. It was just, like, "Okay." (Focus Group, Teachers)*

Specifically, their participation, as indicated in the example above, was seen as a mediator for potential negative effects, including the perception that the program was "something being done to teachers." Moreover, teacher participation was instrumental for creating a shared understanding of both the program and the process for rewarding teachers. For example, a teacher shared, "It's something we might as well be involved in and why not be a part of writing the *process instead of having it implemented on us. So* let's have a part in making, and the staff agreed." A principal echoed the value of teacher's input in the planning:

> *Everybody had a chance and the fact that they all got to help design it. They were the architects of it. So, if they didn't think it was fair, they really couldn't blame [the principal], or the state or the legislature. They had to say, "Well, we created it."*

Another focus group of teachers shared how participation created a common understanding.

> *I think it's not painful because everyone understood what they were supposed to be doing, and they understood the parameters on what we wanted it to look like, and with everyone's input, I think everybody understood. [Teacher Focus Group]*

Benefits of participation in planning

Most teachers reported increased clarification around expectations, including expectations for self-improvement, expectations for collaboration, and expectations for improved student achievement. The collaboration around goals and expectations for achieving those goals was carried through to final reviews in the case of the teachers in the above instance. That is, rather than completing final reviews with teachers individually, a school did team reviews to be consistent with their emphasis on collaboration. For example, teachers explained:

> *When we came in for our final evaluation, we came in as a team this year. In the past, our final evaluation with the principal has always been one-on-one, and we came in as a team, which I've heard positive – nothing but positive comments about that, that [the principal], as well – and it was just a very positive experience. Nobody felt threatened. I mean we all knew where we stood, and it was just a very nice experience. (Focus Group, Teachers)*

Another group of teachers indicated that the program's benefits included self-reflection, as well as collective reflection.

> *It made us look, as a team, on what we were weak on, and it made us realize what we were strong in, and so we could work on the weak part.*

Other teachers noted the gains in collaboration during this initiative:

> *Merit pay has really drawn us towards what we're doing this year as far as more inclusionary efforts because now we can see when everyone works together we actually do make greater gains. (Focus Group, Teachers)*

The *filtering process* used by many schools to discern the elements of their program helped generate engagement, understanding, and buy-in. A teacher further explained this process:

> *The actual process to get to that point, it was really an interesting process where we were presented with the idea, we were to think about the idea, and we were also given time to research, ourselves, about the idea, and then come to the table with the things that we had learned, the ideas that we had gained as we did our personal research, and then we collaborated as a group after our personal – or our little committee collaboration. Then it was presented to everybody, where everybody was able to give input. So it really was quite a filtering process, and people were able to talk about the pros and the cons of each different aspect and weigh in on whether or not they felt like this particular observation was fair and was observable and was even.*

"Costs" of participation in planning

Teachers and administrators who were engaged in the planning process also reported the investment of time, which they indicated was used to preserve communication and collaboration while avoiding a competitive environment. For instance, teachers in one school described the process:

> *And we – those committees met weekly for a while, and then biweekly, and presented it to the faculty. Then went back and hashed it out, so that we had faculty input and teacher input on every single thing. So – and we decided that it was not going to be a competition.*
>
> *It was also individual. Where am I, and what can I do better? And so this was how we came up with the plans.*

The time necessary for planning wasn't constrained to the planning year. For instance, teachers remarked additional planning time also was a precursor to more time spent improving their teaching:

> *I don't think it's just the absolute work that you saw the team meet these indicators that you're doing, but I think it's also the extra work that went into improving our teaching. Okay, what can I do different? Because that took hours and hours and hours, but it's hard to weigh that because I'm thinking, "Okay, I need to change this. What do I do?" So you sit down and you evaluate, and you come up with a new plan, and then you come up with new materials, and all of that takes time. It's not just the essay. It's not just the meetings. It's not just this. It's also what we did to improve our teaching or our team.*

It was not just the time spent in planning committees that was difficult for teachers. Teachers reported that the work and the potential ramifications of participation were also stressful at times. For example, the focus group of teachers below explains the weight of their participation.

> *And it was tense and intense, and I remember being in tears at least twice because I cared so much about how this was going to impact our school and our state and our profession*
>
> *… So it was very tense. I mean we certainly, we talked plenty, but it was also tense I think because, not just because of the issue, but because there were people who had signed on to doing this who didn't really realize when push came to shove what it was going to really be. (Focus Group, Teachers)*

Teachers also described the difficulties and investment in the planning process, particularly assuring that the process was aligned with the goals of the school and the overall intent of the pilot program.

> *A difficulty we faced in the planning process was that we went forwards and backwards a lot. We were trying to set up something that was attainable for everyone, but also personalized so that it wasn't one generic formulated thing that everyone had to do. But then, also something that was going to be meaningful and hopefully not too aside from what you're already doing and teaching. So having it meet all those things and deciding what would be an effective way to do that was a lot of the difficulty we faced . (Focus Group, Teachers)*

Parents also discussed the intensity of the demand placed on teachers. In addition to the intensity of the work, parents highlighted how difficult it was for teachers to reconcile the requirement to be compared to someone else.

> *The teachers really struggle with the notion of having this sort of a program in place because it's a complete paradigm shift from what they're used to, and the notion of being singled out among their peers as being better and deserving more money, most of the teachers – almost all of them hated the notion of that being implemented because they're so uncomfortable with the idea of being compared against each other.*

Reportedly, there were numerous challenges to communication, consensus, and consent when participation in the process was either limited or viewed as limited. Here a principal describes how their approach of using team leaders as representatives of the full faculty sometimes impeded their work due to information not being as salient for others who were not part of the original conversations.

> *[The planning team] could have these great discussions as teams, and these were called team leaders. One teacher from each grade level, so we'd have seven teachers there, kindergarten through six…. [The planning team] could come to consensus there as the team leaders together. This one teacher tries to go back to two teachers that have not been there and explain it, and all the sudden, it sounds different, and they're asking questions, and [the planning team] doesn't know how to answer it. Or they start swaying towards, "Well, yeah, maybe that's right. Maybe this – is this going to divide us?" …And so that was a challenge to take it from this meeting that had had a whole day in discussing this, to take it back to two teachers that hadn't been there and explain it.*

The Need for Support during the Planning Process

Repeatedly, however, participants noted the need for additional assistance and support to develop and implement the program. In particular, there were stated needs for further assistance with defining how to determine each measure—quality instruction, student achievement and parent satisfaction. For planning purposes, schools accessed additional consultants and local support (e.g., USOE leadership; district-level administrative support; Dean Robert Shaw at Westminster; UEPC; charter support personnel). A principal described the value of having an outsider help facilitate the process:

> *The teachers can see [the consultant's] interest in this is to develop the plan, and there's no hidden agenda or anything. Even though I [as principal] couldn't have had one, sometimes teachers think that you may.*

A consistent message throughout the interviews was the need for additional assistance in planning for the details of the plan. This included assistance determining what indicators to choose, how to select and monitor these indicators, and how to have similar indicators across teacher groups (e.g., specialty teachers, other content based teachers). For instance, teachers commented:

> *I think we have no idea what a good range of improvement is for students based off CRT. Like, what percentage is a good percentage to improve? And so that was really difficult for all of us because – and there was really no research or study to really say what was a good range. And so I think that was hard, especially basing off CRT goals is it would be good information for everyone to know what is a realistic improvement.*

Planning for Differentiation and Focus on the Individual

Teachers, principals and other planners alike acknowledged that the expectation that not every teacher would be rewarded or rewarded at the same rate of others was viewed as problematic. In particular, teachers expressed concerns with a program that did not allow for all teachers to do well (e.g., meet expectations or improve). A principal explains:

> *And there's so much progress that you can see made, and it's probably like with the kids, too, that others wouldn't make. So you think, "Yeah, they really deserve this money," but it's just hard to have someone looking over your shoulder when you're doing this and saying the success of this will depend on – the money is differentiated.*

Teachers in a focus group added:

> *Well, that makes it competition among your coworkers when really, you just want everyone to succeed. I wasn't here for that, but that sounds crazy. I would have a problem with that, too….It was very disheartening.*

Although there were feelings of distress surrounding the requirement of differentiation and the measurement performance at the level of the individual teachers, there were also some benefits from the focus on individuals. A focus group of teachers discussed the benefits of the focus on individual goals for improvement.

> *The part that I personally appreciated was the chance for me to decide for myself what I needed to improve, what I wanted to improve, and how – the how part was not as clear or it didn't feel like it was as much my choice as much as just what I wanted to do, and I was really appreciative of that because there's a list this long of things I want to do. So it was nice to just pick one of those and kind of zero in on that and then have a measuring tool, whatever that may have been, for me to just say, "Oh good. Now I'm really good at this thing," rather than usually I'm trying to do all nine or ten things at once and then I'm not as effective.*

Teachers also discussed the positive attributes of a performance-based compensation pilot program that permitted them to have their own individual goals.

> *I just think about that was like both good and challenging was that we all had goals that we had to meet in certain general areas, but how to meet those was left up to us individually, which was a good thing in one way because it's respecting who we are individually and we can set our own goals.*

Another teacher discussed specifically the increased focus on their own improved instruction:

> *I think the whole thing kind of provided some insight into what I'm doing, and tried to say, "Okay, are there some areas that I could do better at?" And it gave me some encouragement to try to improve in those areas. And so I think it was positive from that standpoint.*

In general, principals indicated that the program has permitted higher degrees of self-reflection.

> *I don't see anybody saying that we're all good teachers and we don't need this, where as that's where we were before. It was almost like, "Let's all just be nice to each other and leave things as they are," and now we're just looking at, "Let's share with each other. I want to see what you're doing, and let's talk about this." That's what I'm hearing now instead of, "Oh, we're all good, so why don't we just divide the money and give some of it to each of us?" So I think it's a lot more meaningful, and I think we're thinking of new ideas all the time and open to new ideas instead of thinking, "Hey, I've got my plan." (Principal)*

Repeatedly, there was acknowledgment for how the participation in the program served to enhance self-improvement.

> *For me, I thought it was hugely helpful, because it was something that came from me that I felt like I needed to work on. So it wasn't something that I was given, saying, "Here, work on this" when it didn't apply to me.*
>
> *I was able to pick something that I felt I needed to get better at, and perfect it.*

Planning for Collaboration and Focus on the Group

A recurrent concern was the impact of a "competitive" program. Specifically, participants indicated that efforts to use programs that situated teachers "against one another" were viewed as lethal. A principal explained:

> *I think that if they ever go to teacher versus teacher, it'll really destroy public education. I think if you keep it as a professional or a community model, and the fourth grade, if one goes down the whole group goes down, that's powerful.*

While an individual-based system was seen as positive, there were recognized shortcomings. A teacher explained the difficulty in having a system that built in individual goals but then compared teachers to each other.

> *It was very difficult for somebody like me, who wants clear parameters….It was hard because it wasn't consistent. What I was working on might be totally different from somebody in my grade level, which is okay, but how do you measure that? I think that was the hard thing is how can you put this is equal to this, this is better than this, this is higher than that.*

The principal added that there were benefits of maintaining a system which coupled accountability with potential for individual remediation.

Collaboration among teachers was repeatedly addressed by teachers, principals, and other stakeholders. The maintenance of collegiality and collaboration was perceived as a necessary element to successful implementation. Teachers explained their intent to maintain collegiality and avoid competitiveness among their school's faculty. For instance,

> *Well, you really had to depend on your team members to do their part. Whatever the assignment was, you depended on them to pull through; otherwise, we were going to fail, so – and everybody stepped up to that, and I think that was a huge part of the development of our program because we were not going to be competitive. We all said, "If this is going to be competitive and we're going to be eating each other alive, then we're not going to do this," and that was at the very, very beginning, and we carried that through. (Focus Group, Teachers)*

The need to create a non-competitive environment was echoed across teacher focus groups. For instance, teachers explained the role of collaboration and noncompetition held in their participation in the pilot:

> *That seemed to have worked for us, and not pit teacher against teacher, and it helps us to evaluate ourselves, our students, our whole school, and we own every child here, and the problems of the school. It was a better approach. (Focus Group, Teachers)*

Another group of teachers explained the cautions of not creating a collaborative process. *"We worked as a team. But if other schools didn't have that, you could see where it would really shut those doors and divide them even more."*

Parents also remarked on the nature and existence of collaboration among faculty, particularly the observation that there was more collaboration and that those who had collaborated before were putting in more effort.

> *I think that it made collaboration maybe a little bit stronger in some of those grades that might've been weak. I think that they might've seen the benefit of collaborating more together, and so they – Well, and some of them even took it to kind of the next level. Even the ones that were strong. Because some of them – right, the ones that were strong went into the – you know, in the group planning of lessons and more of that. (Focus Group, Teachers)*

A Tiered Approach: Focusing on the Individual and the Group

There were several types of plans developed, including those that rewarded just individuals and those that rewarded individuals based on individual, team, and/or school award. The variant type of rewarding an individual based on the compilation of their goal attainment and their team and/or school goals was viewed positively. For example, a principal explained why she preferred a model that rewarded teachers based on what the school, team, and individual contributed.

> *The thing that appealed to me is that it still held the individual accountable to a certain degree, but it also promoted collection at the team and school level. I want my teachers not only to collaborate with one another on a first grade level, but on a second grade – from first to second. So I could see that it was a structure that would support collaboration, and I really liked that because that was my very first fear when I introduced it to the – or not my fear, but the staff was "We are not going to do anything that's make us be competitive."*

A district administrator added that the tiered model for performance-based compensation worked ideally to promote collaboration.

> *This [the school's] model is a really good model because of the way it's structured with points where you get team points, you get school points, then you get your own points. So it actually could, if you do it careful enough, could promote collaboration. And that was the whole philosophy behind the model was to somehow promotion collaboration and still have performance pay, which is like oil and water, but you can do it.*

Implementation

Below is a discussion of the key themes regarding implementation of the actual plans, including the role of the principal, changes to professional practice, feelings of uncertainty about whether the benefits were worth the effort, and appreciation for being recognized for their efforts.

The principal played a key role in implementation

In addition to the principal serving as a key player in the development for each school's plan, the implementation of the performance-based compensation pilot program was heavily reliant on the work of the school principal. The principal also served as a monitor, appraiser, and administrator of the bonus calculations. The extensive involvement of the principal was recognized by teachers and other stakeholders alike, including recognition of the time and effort. As a focus group participant shared, *"[The principal] definitely was the one that took on the brunt – all the brunt of it.*" In other instances, principals were called the "backbone" of the initiative for the role in implementation. One principal explained the necessary effort and attention of leadership's time in ensuring effective implementation.

> *I think there were more – I think it was surprising how it pushed me as a person both physically, emotionally, and mentally. It was really draining on me. This is something*

> *you think "it's going to help student achievement." To lead this as a principal and to do it in a way that I felt had integrity took everything I had. And then also, I think I was pleasantly surprised that it brought us, I felt like, more together.*

Another principal explained in detail the time necessary to fully implement the program in ways that are beneficial:

> *I think that if you're just collecting data or doing something as a checklist for what somebody wants you to do, in this Performance Based Pay, it's not as meaningful.*
>
> *While there are certainly benefits of having the funds determined by the principal, it is also a time consuming process. Additional district or state support, however, will be necessary to do this long-term as insufficient time is available at the school to do this long-term.*
>
> *if there's systems in which it's happening naturally or even if it's something new that there's a expectation it's going to be not just this one-time thing and then it becomes a little bit more routine and they get the district, whoever, even if it's a special person on your staff but somebody that can take on – because it does take a little bit of extra work to compile data, collect data.*

The process of implementation itself altered professional practice

Principals discussed how the participation in the performance-based compensation program altered their leadership practice. These changes included modifying structures, use of principal time, and questions that were considered regularly. A principal explained:

> *I had to change the way I did my job in order to make my instructional leadership happen. That was huge. I had to completely revamp the organizational structure. I had to turn over my calendar to my secretary. I had to teach her how to screen my calls, how to – I had to learn to let go of things.*

Frequently, participants reported the generation and use of questions that were used to guide practice. In many instances, principals and teachers discussed how the performance-based compensation pilot program provided an impetus for teachers to take risks on new methods.

> *"Yeah, let's refocus. Let's take some risks here. Let's experiment with," How can we influence and how can we impact our test scores – that's the bottom line. But also how can we impact our craft, how well we do our craft? Because if we can impact our craft, we can, as a result, impact our test scores. (Principal)*

As a facilitator in the implementation process, another principal discussed how they worked through questions during implementation.

> *So what I wanted to do as a leader was I wanted to lead this and facilitate it, but I want the discussions to be that the teachers had to do the work. They had to think about it. They had to look at the ideas. They had to say how do we want to make this work for us. So doing it that way takes a lot more time, and you go through a lot more emotions, but I think in the end, it can be more effective. (Principal)*

Uncertainty as to whether the effort required was worth the bonus and what will happen now

Principals and teachers generally reported the planning process and the implementation of the school's individual plan as beneficial. However, there were concerns as to whether the available resources for rewarding teachers were sufficient to serve as a motivator. The questions posed now that the pilot has

expired also indicate reflection on whether or not the program was "worth it." For example, a principal noted, *"again, are my teachers' time worth what we did time wise? Will they keep wanting to participate at the level they did without the bonus."* Teachers also commented on the time necessary to implement the program. The following responses are representative of what teachers had to say about the time they spent on the requirements of these programs:

> *Much more paperwork. I think it took time away from the students. I was always worried I didn't have enough data. It was a lot of extra work.*
>
> *I strongly disliked the extra time I had to stay after school. I felt like I could have used it to plan lessons. (Focus Group, Teachers)*

In the end, many expressed that they did not believe that *"the money was not worth the time and work*." *(Focus Group, Teachers).* Another question is whether changes in practice achieved during the pilot program can be sustained in the absence of a bonus. For instance, a principal noted, *"The only thing I worry about is the amount of money a small school can come up with because I think as an administrator I found out it's valuable."*

Teachers appreciated the bonus as recognition for work well done

At the same time as the total compensation was questioned, participants did indicate that the recognition was valued, including the compensation and acknowledgement from others. Teachers indicated that the recognition for participation and a job well done was appreciated.

> *I think it was positive just kind of on a celebration standpoint, and I may be speaking out of turn here, because I haven't – I don't know, individually. But I kind of got the feeling that even if you didn't earn as much money, we were all semi-satisfied with, like, wow, look at all the good things that people are noticing that I'm doing. (Focus Group, Teachers)*

The additional pay to teachers was perceived as a form of recognition for the work that they did.

> *I agree, you can't really put a dollar value on the work you're doing with kids, but it was – I mean I didn't do those things for the pay, but it was nice, you know, to receive that little like nod of, "You worked hard. Here you go."*

The recognition for their work was described as motivating. In the surveys, teachers also expressed that the financial reward was perceived as for a job well done.

> *"At my school, the teachers are extremely dedicated and motivated and this pay was just a reward for a job always done well."*
>
> *"I saw a reward finally for the hard work I go through every day to teach students."*

According to some teachers, they often felt like there time could have been better spent and the general sentiment was that the money, although nice, did not sufficiently cover the extra work that they had to do.

Student Achievement

Although we did not find that the pilot program had an effect on student achievement per se, we did gather important information about the school and instructional practices related to student achievement. Specifically, we found the following key themes related to student achievement during our interview and focus group discussions:

- Use multiple indicators (more than just CRT score to measure student achievement), including formative assessments.
- Increased focus on data related to student performance to inform instruction.
- Consideration of impact of student achievement measures.

Use of Multiple Indicators and Formative Assessments

Importantly, the five participating schools did not solely use CRT scores or other "one score" measures. Instead, they used a compilation of indicators. To this end, the schools did focus on use of other assessment data. A principal described their school's use of assessments to both inform the quality of instruction and make instructional decisions.

> *So there's this whole district philosophy that we are heavily engrained, heavily invested in assessment. But the kind of assessment that we're heavily engrained in, is the way it should be done. It's not just a standardized benchmark test. It's progress monitoring that informs and substantiates the benchmark test.*
>
> *And so our staff is very familiar with, and they're very comfortable with the testing, assessment, and accountability, not just to inform their quality of instruction – that's important – but they use it heavily to guide their instructional decisions about children and how do we differentiate for children. So, they are actually informing their instruction based on this assessment. (Principal)*

Schools reported that students' overall learning improved during the course of the performance-based compensation pilot program. In some instances, they noted that learning gains were visible in the areas of focus (e.g., math). They did question whether gains were made across subject matter. Importantly, across all participant groups, there was consensus that one score is insufficient to evaluate the effectiveness of an educator. For instance, the following is representative of comments made by participants regarding the determination of what measures are used in the performance-based compensation pilot program plans.

> *Something right off I guess I would just hope that the people that are making the policy and the laws would recognize that teaching is a multifaceted job and that one score, whether the CRT score or something, is not a measure of the whole thing but if we were gonna have one score, ever, I would hope that it would be related to some kind of a progress score or value-added score rather than a proficiency score or if they do ever do proficiency that it's over time, following students, of course, but that then becomes a progress score if you do that.*

Another participant discussed the concern related to centralizing test scores:

> *That's kind of what's disturbing to me is when you read all these articles and the comments that people make that are not in our district is the only thing they talk about when you say performance pay is test scores. That's the only thing they talk about, and ours wasn't like that at all. The scores were important and they always have been, but it wasn't the only thing. (Focus Group, Teachers)*

Parents also commented that they perceived an expansion in the use of assessments throughout the schools. They noted recognition of how often teachers use multiple assessments now.

> *I'm learning that the teachers are always assessing the children in different ways. It's just not always written on paper, but in how they work on their own, how they verbalize,*

express themselves. Teachers are always assessing the child's performance in more than just on a test, on a written test. (Focus Group, Parent)

There was discussion throughout the teacher focus groups and principal interviews regarding the increased development and use of formative assessments. A principal explained the rationale for having ongoing formative assessments:

So timely, access so we can make fast judgments. We can't wait for three months to get the results back. We have to have them the second the test is done. And we changed our scheduling and we're very flexible and very fluid with the data that we have.

While the expansion of (formative) assessments was viewed positively by teachers, they also recognized that this could have a detrimental side as well. Teachers explained:

A ton of testing, a ton of quizzing, and towards the end. It felt like I was getting poorer results because my kids were so over testing that it just wasn't worth what I was putting into it towards the end. (Focus Group, Teachers)

Increased Focus on Data

The utilization of more formative assessments reportedly led to increased use of data to inform instruction. This observation was noted by teachers, principals, and parents. In fact, parents indicated more teachers "re-*teaching that concept until is understood*." Teachers and administrators concurred that there was an increased use of data to inform instruction.

I saw a major increase of looking and implementing what data told you at this school. And I don't think that would have happened without the pilot….[The teachers] have become much more data-driven in the last year, in my opinion, and much more looking at – they've kind of had a shift in looking at what students are learning versus what they're teaching, which is great. [District Administrator]

Teachers also noted the increased use of data as a positive change:

The comment I was going to make is it really made us concentrate on data more than we ever have before, and that was what our rubric concentrated on, as well. So we were looking at pretests and posttests, and so our PLC's really focused on data where the students were succeeding, which we had done in the past, but even more so, I think, this year and concentrated our PLCs on where we could help those students and what successes we were having.

Teachers discussed how the data prompted attention to progress in student learning of the core content in preparation for the CRTs:

I think the most important data point was that mid-year assessment. Because when we took that, we said, "Oh, there's some kids, they're not going to make it if we don't do something different," and it was good indicator on how they would do on the CRTs, as well.

Instructional Quality

Instructional quality, which was a primary focus of the performance-based compensation program, received considerable attention at pilot schools. This attention was by principals and teachers alike.

Peer and principal observation of classroom teaching was extensively used as a tool to improve instructional quality and was considered beneficial in several ways

One area of increased activity was observations of teachers. In some instances, this included observations by peers, school administrators, and occasionally district administrators. Multiple instruments were used, including Robert Marzano's and Charlotte Danielson's iObservation®, and school-developed instruments. In several cases, these increased observations are continuing this year, including videotaping lessons and peer or collegial review of the video observations with structured conversations. Benefits of these observations included sharing resources and talents, reviewing data, modification to instructional strategies, collaboration, relationship building, and development of stronger trust among staff. Importantly, there were requests for further professional development on how to do observations in meaningful ways, as well as what to do with the information gathered from them. Below teachers describe some of the benefits of peer observation:

> *Styles are different. You can be teaching the exact same thing, but with a little bit different style you learn so much from just a few minutes of seeing someone else.*
>
> *It just opens your mind so much more than when you're in your own classroom doing your own style and your own things all the time. (Focus Group, Teachers)*

Observations, though not always convenient, were seen largely as beneficial. A focus group of teachers indicated:

> *I think iobservations, I hate them. But I think they're really important. I do. I think they keep you on your toes….*
>
> *Any time you've got another person there, I think you do better.*
>
> *And you know it's not just little people that are watching you. You know, you're more aware of what you're doing and making sure that you're hitting all the things that you're supposed to do to be a good teacher and have a good lesson. (Focus Group, Teachers)*

Teachers frequently discussed the observations of the principal and their focus on instructional practices. In this case, the principal had conducted two observations and provided feedback. The teachers elaborated on the growing acceptance of the principals' observations and the value of these observations:

> *The first time I didn't pay a lot of attention to it. I just said, "Oh, [the principal] is not here every day. [The principal] doesn't know." But when [the principal] caught it the second time, I thought, "Okay." So you know, I appreciated that because then I thought, "Yeah, I do need to work on that.*

In general, teachers noted that the constructive feedback from principals was worthwhile.

> *And she's really positive so that you can talk with her after and she'll always say, "I noticed that you did this well," or, "I noticed this. Let's try it and I'll come and watch next week and we'll see how it goes," and you try – you know, she gives you a suggestion that you can try and so that improves my teaching is her positive attitude, the way that she approaches – not criticizing but, you know, giving you a constructive criticism, yeah, to become a better teacher.*

For their part, principals also indicated that the observations have helped them focus on instructional quality as well as student achievement. Moreover, the observations helped the schools focus attention on instructionally-based conversations. For instance,

> *Well, you asked me about the – when are we having these conversations, and I said, "All the time everywhere." When I started coming to the classroom, I told them the first year, "I'm coming because I want to know if I can see it. I'm coming to train my eyes. I'm not necessarily coming to make an evaluative judgment of you. However, I will be sending you feedback that's going to feel judgmental, and it's going to feel like I'm evaluating. But the purpose is I want you to tell me if I'm nailing it. Am I seeing it?"*

In a few instances, principals utilized video-taping to conduct observations and debrief with teachers regarding their instruction. These video-taped lessons were viewed as beneficial by both the school's leadership and teachers. A principal described the videotaping in this way:

> *Well, at first teachers would say, "Oh, I don't want to be videotaped." But then when they sit down and watch them, what I would do is so incredible because I have it on tape, the feedback I give them while we're watching. And so I can say to them – and I videotape the kids a lot, and they love it because they never really get to see their kids from that perspective. So it's not just camera on them. There's a lot of me zooming in on the kids, too, so they can see this is what the kids are experiencing. This is what you're doing, and I didn't have to say a lot. They would pick it up. I had no idea I picked up the pencil and did the work for that child at that time.*

> *I tell my kids not to do that. I can't believe I did that. Or now I'm really moving slow on this lesson. I'm even getting bored on it. But the thing that I could do is just mention about how do you think that child felt when you just said that, or look at your presence there with that class, or mention the things that they could leave and just go, "Wow, I am a good teacher." (Principal)*

Trust among the faculty was essential in obtaining the benefits of observation as a strategy for instructional improvement

Teachers talked about the trust that needs to be developed and the imminent concerns related to building trust in peer observations. For instance, a teacher in a focus group explained:

> *We work together so much, and we're used to the principal coming in and out, but to have the peers come in and observe us, that took just a little bit of getting used to. But then I think we built an even bigger trust.*

This was extended to how teachers and principals discussed the freedom to modify instructional practices. As noted earlier, developing a positive school climate with high degrees of trust was seen as an imperative for successfully implementing this performance-based compensation plan. In this instance, a principal explains how trust and a positive school climate mitigated what could have become a toxic culture.

> *I think it would have to have a really positive school climate. I think we would have to have a school climate where there was a climate of trust. It was a place where people enjoyed coming to work. It was a place where people talked to each other and cared about each other and looked at each other as real people and knew each other. I think all of those human being traits really help in something like this because if you didn't have those things and then you add a monetary value to things, you could have it just – it would be very toxic.[Principal]*

A good system was one that both built trust and built upon it. This created an atmosphere that promoted sharing among colleagues, self-reflections, and improved instruction.

> *They need process that promotes reflection and trust, and if you can build trust amongst the team to then be self-reflective, to say, "You know what? I could've done that a little bit better." That's when you get serious growth from your instruction and your students' achievement.*

Finally, the necessary role of establishing rapport and trust was evident even among those not in the schools on a day-to-day basis. A district administrator explained how important rapport was when a principal, or others, are going to be observing teachers frequently:

> *[The principal] made a major focus on rapport. "I've got to get my teachers comfortable with me or none of this will work." And she's really made that a focus in [the school] administration, and it's paying off.*

Parent Satisfaction

We found a number of key factors related to parent satisfaction and the strategies for measuring satisfaction in relation to school and teacher performance.

- Producing a valid measure of parent satisfaction was challenging.
- The rationale for and benefit of using parent satisfaction as a measure was recognized.
- Communication with parents increased generally.

When asked how one determines a good teacher, parents listed such attributes as caring, classroom control, classroom management, structure and control, happiness, engagement of children, and differentiated instruction.

> *You know, are they engaging the kids in different activities? And if this activity didn't work for Tommy over here, did we do something different and try – you know, because not everybody learns the same. So did we try something different, or is she just standing up there? I mean, I can go to the junior high and tell you who's not a good teacher over there. You know, there's a few that have been doing the same thing every year. They stand up and lecture, they fill the worksheet out, they put a video on. That, to me, is not a good teacher. You know, but are we doing different things? Are we, you know, trying to reach out to every child and their different learning abilities? (Focus Group, Parent)*

Producing a valid measure of parent satisfaction was challenging.

Principals and teachers alike noted the challenges regarding the measures of parent satisfaction. Schools administered surveys to determine degrees of parent satisfaction. Of particular concern to principals and teachers were issues regarding parents' first-hand knowledge of their teaching practices or effectiveness, or how parents would use their "child's version of events" or other parents as the sources of information for their judgments. In addition, there were issues regarding response rates. A teacher expressed the common concern over what to ask parents and how well parents were in a position to answer questions about the quality of instruction or student achievement:

> *How valid is it? Does the parent really understand what's going on there? And I understand that we're in the business of perceptions, and that's important to know, but you know, you also have to look at some of those and say, "Okay, here's obviously someone who's had a bad experience that is tainting other perceptions." And you know, you – I don't know, that's always a challenge, the parent survey, because are the questions worded in a way that the parent understands what – you know, and it tells you what you think?*

Another focus group of teachers reiterated concerns about the types of questions asked of parents:

> *There was one question that said something about – remember that one about instruction? And we felt that the parents weren't, couldn't judge us on the quality of our instruction because they're never in the classroom*

The rationale for and benefit of using parent satisfaction as a measure were recognized

Again, despite concerns over response rates and knowledge of classroom practices, there were noted benefits and reasons cited for a focus on parent satisfaction. Teachers discussed how parent satisfaction was important, even within the larger scope of the system.

> *I think it's okay to weight teacher quality and the student performance equally, and I think there needs to be a place for parent satisfaction. That's a big part. I mean, especially in today's education world where we're getting more options for parents, I think we need to be aware that there is some PR involved, and that didn't use to be the case, necessarily, but it is. We need to interact with our community more and our parents more. And so, that was good for us to be aware that we're representative of public education in the elementary, and we need to show what we're doing and, kind of, brag about what we do a little bit.*

Another group of teachers explained how the feedback generated from the surveys and principal observations was helpful to guiding their practice.

> *I think it was still hard for people to attach a dollar sign to the actual work that teachers do because how can you do that. Just philosophically, I guess, it still is, I think for faculty kind of hard to do that. But it was motivational to have feedback. I mean because you'll definitely become better with feedback from your students and your principal and community*

For their part, the parents interviewed expressed also mixed feelings about the use of parent satisfaction for determining teacher bonuses. For instance, a parent explained how important it is for parents to be able to contribute feedback:

> *I should have some say. … you can tell, I mean even if you're not in the classroom you can tell if your child is happy and learning and progressing when they come home, don't you think? I mean you should be able to say, "Yes, that teacher's a positive force in my child's life and they're making a difference." I mean you should be able to say that. I mean and if they're not you should be able to say that also.*

This perspective does reiterate the teachers' concern regarding the issue of parents in classrooms. However, it also emphasizes the value of parents getting information from their child or making judgments regarding their child's learning and experience. This view was supported. For instance, a district administrator commented:

> *If every parent in our community – if he gets negative feedback from the parents, that's something he's going to have to work on. If he's got bad rapport with students and parents, that's a key to his job. It's not as big a key as data and just flat out instruction, but it's important.*

Teachers increased their efforts — in depth of information, frequency, and variety of methods — to communicate with parents, but were not always successful

Teachers reported more communication between school and home during implementation of the pilot program. In particular, this increase in communication was attributed to the consideration of parent

satisfaction. They also indicated that their practices reflected additional time investment in what kind of information was collected and distributed for their communications with parents. For example, teachers in a focus group explained:

> *I do feel more aware of what I let parents know and how I – but I always think I – I think I always was trying to let parents know in any way that you can what their child needs to work on and how they can help and when they-*
>
> *But I think it was more, like you say, because I had to report I was more tuned in, "Oh, okay, I better go write that down so I can report to Johnny's mom that thus-and-such happened this week." But when I did that then something else had to move out of my space for me to accomplish that. (Focus Group, Teachers)*

Another teacher expanded on the types of communication used last year, which included updates on student progress:

> *I really focused on my parent communication more last year than I probably have ever had. Like I did a lot more making parents aware of where their students were; how they were succeeding and not that I didn't do it before, but I just was more aware of it and more conscientious, I guess and just made sure that I was doing my part, letting the parents know what was going on in the classroom and even with their individual child in my classroom.*

Parents' comments also reflected the perceptions about communication. For example, a parent commented on the changes seen during the implementation of the pay for performance:

> *Definitely the communication between teacher and parent has improved. I saw a drastic change that year that I was getting more feedback of what was happening in the classroom, and very specific areas, exactly where my child stood. Even within two weeks after that, even an update on this is what she needed help with, and now this is where she is. And it was so immediate, and I'd never noticed that before in any – I give that attribute to the teacher is really trying to work with each student at their level and trying to bring them up a level, especially in math.*

Another parent remarked that during the pilot year, teachers were focused on instruction and student learning.

> *I feel like the teachers, you know, were kind of on their A-game. I mean, they're always good, but I mean, I felt like there was a lot of focus on that. We had a day where we came and listened to each grade level talk about kind of what they're doing and how they're collaborating, and were able to get a real feel of what was going on in each grade level.*

As schools focused on particular goals, this information was communicated to parents as well. A principal noted how they were using results from the first year survey to modify how the school communicates with parents. She indicated that the perception was that while the school thought they were doing a great job with communication, the parents were *"not feeling it."* The principal added:

> *So, we have to do something here to work on this." So everybody's going to have a newsletter on the website. Everybody's going to really work on getting parent volunteers into their classroom. Everybody is going to work on eye contact in the hallway and we're going to have badges that say, "Visitors –" It's not about safety. It's about, "Well, this is a stranger in our building and do we need help?"*

CONSIDERATIONS

Leadership noted the need for a system that permits differentiation to recognize those teachers whose instructional quality is superior and those who are effective but perhaps not superior. Teachers also indicated that a program that recognized differences between teachers was warranted, if it could be implemented in a way that did not breed competition. Of concern was the issue of developing a system that could couple accountability with potential for individual support, and where when needed remediation and additional support could be provided for teachers, teams, and schools to meet expectations.

Teachers added that an emphasis on improving teaching and instruction would serve the performance program well. For example, a teacher explained:

> *If their intention for performance pay is to get rid of bad teachers, it will fail. If their intention is to get rid of bad teaching, improve instruction, and tie it around PLCs, then it could be successful.*
>
> *Sill do the three categories, that it might be helpful to have a menu under each one and rather than having a total blank sheet of paper, have – have multiple ideas and, say, "Maybe you're not limited to this but here are ways that have been tried other states, other schools and pick – almost like pick two from this category or one off this list and make up one of your own," or, you know, something like that. Then I don't think that it would be as hard but – but then you might skip over some of the interesting discussions, too.(Focus Group, Teachers)*

According to participants, it is also imperative that additional clarity is needed to ensure that pay-for-performance programs not be about "firing bad teachers" but instead about rewarding "good" teachers. The perception that the underlying intent of the program is to "weed out teachers" was prevalent. Participants indicated that teachers generally were working hard and dedicated to both teaching and students. As a result, many found the idea of rewarding teachers in this way wasn't warranted, and in some cases "insulting"

> *So it was – the other part of it for me was it didn't feel like being treated like a professional, because, you know, it was somebody assuming that you were holding back, and if you just had – if you were just given that extra $2,000.00 you'd produce better teachers. (Focus Group, Teachers)*
>
> *I don't mind being offered money, because I'm doing it for a living. But to be offered money in the sort of assuming that I'm already not doing the best I can is insulting. And especially since all of us go home so exhausted that we don't have a single brain cell left. (Focus Group, Teachers)*

The Performance-Based Compensation Program was approved and funded by the Utah Legislature for a planning year plus an implementation year. The five schools that participated in Utah's Performance-Based Compensation Program—Ashman, Canyon Rim, Manila, Midway, and Wasatch Peak Elementary Schools—were intricately involved throughout the planning and implementation process. From their experience and the analysis of the implementation process and associated outcomes, we provide the following considerations for future planning and development of a strategic compensation program.

Pre-Implementation

- Plan for sustainability (e.g. sufficient resources, including amount and longevity), including allocation of funding for multiple years to support a planning year and at least three years of implementation.

- Provide guidance for implementation, including general framework (e.g., instructional quality, student achievement, and student, parent, and community satisfaction) for the program.
- Consider a menu of potential measures for schools and districts to select from for instructional quality, student achievement, and student, parent, and community satisfaction.
- Ensure availability of program to diverse schools.
- Establish a network of program participants with regular communication, dissemination of information and success, and opportunities to convene participants, including administrators and teachers.

School and District Planning

- Provide leadership for the vision for school and definition of the purpose of the school's involvement in the initiative.
- Ensure and/or develop a climate of trust and collaboration.
- Develop a planning team and process that is inclusive and representative of the school staff (e.g., teachers, specialists, paraprofessionals) and community (e.g., parents, school community council members, district administrators).
- Provide compensation for those engaged in development of plan.
- Support for the development of multiple measures appropriate for school context within the instructional quality, student achievement, and student, parent, and community satisfaction.
- Use data to inform selection of measures and indicators for success.
- Provide technical assistance for the development and use of the common assessment tools.
- Develop tiered plan that acknowledge the individual's contribution, a team's contributions, the overall school's progress.
- Utilize external resources (e.g., district and state staff, consultants) to support planning process and development of plan.
- Plan for alignment with other key initiatives.
- Allow schools to create plans that are customized to the individual school but require at least one common assessment for each of the areas of evaluation.
- Communicate consistently with all stakeholders regarding process, timelines, expectations, and anticipated outcomes.
- Expend time and effort to develop buy-in among all stakeholders in the process.
- Use all measures formatively.

Implementation

- Maintain communication consistently with all stakeholders regarding process, timelines, expectations, and anticipated outcomes.
- Focus on instruction and instructional quality.
- Provide support for school leadership to administer the program.
- Use data to utilize the lessons learned from the implementation year.
- Build capacity (e.g., professional development, training, shared learning, networks) to develop, implement, and refine plans.
- Provide professional development to support staff (e.g., improving instructional quality, increasing student achievement, improve student and/or parent satisfaction, development of common and/or formative assessments, use of assessment data, data analysis, structured improvement based on data).
- Monitor implementation for plan improvement and fair application of the plan.
- Utilize annual results for school improvement efforts.

REFERENCES

Barry A. Friedman, Paula E. Bobrowski, Dana Markow, (2007) "Predictors of parents' satisfaction with their children's school", Journal of Educational Administration, Vol. 45 Iss: 3, pp.278

Eberts, R., Hollenbeck, K., and Stone, J. (2002). Teacher Performance Incentives and Student Outcomes. *The Journal of Human Resources*, 37(4), 913-927

Kingdon, G., & Teal, F. (2002). Does Performance Related Pay for Teachers Improve Student Performance? Some Evidence from India. Retrieved from EBSCO*host*.

Mahony, P., Menter, I., Hextall, I. (2004). The emotional impact of performance-related pay on teachers in England. *British Educational Research Journal*, 30(3), 435-456.

McCaffrey, D. F., Pane, J. F., Springer, M. G., Burns, S. F., Haas, A., & Society for Research on Educational Effectiveness, (. (2011). Team Pay for Performance: Experimental Evidence from Round Rock's Project on Incentives in Teaching. *Society for Research on Educational Effectiveness*, Retrieved from EBSCO*host*.

Podgursky, Michael J., and Matthew G. Springer. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management* 26(4). 909-949.

Ritter, G. W., & Jensen, N. C. (2010). The Delicate Task of Developing an Attractive Merit Pay Plan for Teachers. (Cover story). *Phi Delta Kappan*, 91(8), 32-37. Retrieved from EBSCO*host*.

Springer, M. G., Ballou, D., Hamilton, L., Le, V., Lockwood, J. R., McCaffrey, … Stecher, B.M. (2011). Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT). *Society for Research on Educational Effectiveness*, Retrieved from EBSCO*host*.

Wohlstetter, P., Nayfack, M.B., and Mora-Flores, E. (2008). Charter schools and "customer" satisfaction: Lessons from field testing a parent survey. *Journal of School Choice*, Vol. 8, 66-84.

APPENDIX A: TEACHER SURVEY

Note: The exact from of the questionnaire was administered online.

Introduction: This questionnaire is part of an evaluation being conducted by the Utah Educational Policy Center on behalf of the Utah State Office of Education concerning the Pay for Performance program piloted in your school last year. In answering these questions, you will help us to identify the strengths and weaknesses of this program as it was implemented in your school, and the effect that the program had on teachers. We very much appreciate you sharing your responses and opinions with us. Please note that this questionnaire is anonymous—we will not link or attempt to link your responses to your identity in any way.

School:

Program Development

Did you help to develop the Pay for Performance program at your school? Yes No

(Route based on yes or no)

If yes: what role did you play in the development? (Open text box)

If no: at what point did you realize that there was an incentive program in your school? (Open text box)

Personal Experience

Did you modify your teaching of mathematics in any way as a result of this program?

Yes No NA

If yes: please give specific examples of how you modified your teaching.

(Open text box)

Did you modify your teaching of language arts in any way as a result of this program?

Yes No NA

If yes: please give specific examples of how you modified your teaching.

(Open text box)

Did you modify your interactions with the students in any way as a result of this program?

Yes No NA

If yes: please give specific examples of how you modified your teaching.

(Open text box)

Did you modify your interactions with parents in any way as a result of this program?

Yes No NA

If yes: please give specific examples of how you modified your interactions.

(Open text box)

Suggestions for the Future

In your opinion, what were the positive aspects of this program and how could those positives be accentuated in future incentive pay programs?

(Open text box)

In your opinion, what were the greatest downfalls of the?

(Open text box)

Can you think of any unintended consequences (either positive or negative) related to the implementation of the Pay for Performance program in your school?

(Open text box)

Please signal your level of agreement with the following statements. Any comments will be greatly appreciated.

I understood **how** to receive a reward through my school's incentive program.

Strongly disagree	*Disagree*	*Neutral*	*Agree*	*Strongly agree*
1	*2*	*3*	*4*	*5*

I felt like the distribution of Pay for Performance funds was fair.

Strongly disagree	*Disagree*	*Neutral*	*Agree*	*Strongly agree*
1	*2*	*3*	*4*	*5*

I expected to be rewarded through this program.

Strongly disagree	*Disagree*	*Neutral*	*Agree*	*Strongly agree*
1	*2*	*3*	*4*	*5*

The rewards offered were of sufficient value.

Strongly disagree	*Disagree*	*Neutral*	*Agree*	*Strongly agree*
1	*2*	*3*	*4*	*5*

This program improved the school experience for teachers.

Strongly disagree	*Disagree*	*Neutral*	*Agree*	*Strongly agree*
1	*2*	*3*	*4*	*5*

This program improved the learning experience for students.

Strongly disagree	*Disagree*	*Neutral*	*Agree*	*Strongly agree*
1	*2*	*3*	*4*	*5*

Please use the space below to provide any comments about your responses to the previous items.

(open text box)

Please rank the following items from 1 to 7 in terms of their importance (with 1 being the most important and 7 being the least important) in determining the pay for teachers:

Student academic performance	1	2	3	4	5	6	7
Test scores	1	2	3	4	5	6	7
Teacher's education	1	2	3	4	5	6	7
Teacher's experience	1	2	3	4	5	6	7
Instructional quality	1	2	3	4	5	6	7
Classroom management	1	2	3	4	5	6	7
Parent satisfaction	1	2	3	4	5	6	7

Please list any other factors that you feel should be considered in determining the pay for teachers (open text box)

In your opinion, what were the positive aspects of this program?

(open text box)

How could the positive aspects be accentuated in future incentive pay program?

(open text box)

In your opinion, what were the greatest downfalls of the Pay for Performance program?

(open text box)

Were there any unintended consequences (either positive or negative) related to the implementation of this program in your school

(open text box)

APPENDIX B: PARENT QUESTIONNAIRE

Note: This exact from of the questionnaire was administered online.

This questionnaire is being administered by the Utah Education Policy Center on behalf of the Utah State Office of Education. Your answers will be used to help make our schools better for students and for teachers. Please answer these questions as they relate to your elementary school student's experience at [name of school] *school* ***last year*** *(2010/2011 school year). If you had more than one student in the elementary school, please answer the questions as they relate to your oldest student who attended* [name of school]. *Your responses will be completely anonymous. If you did not have a student in this school last year, please disregard this questionnaire.*

1. What grade was your child in last year?

Kindergarten
First
Second
Third
Fourth
Fifth
Sixth

2. Who was your child's teacher?

(Drop down with names of teachers)

3. What grade would you give your child's school?

- A
- B
- C
- D
- F

4. What grade would you give your child's teacher?

- A
- B
- C
- D
- F

5. On a scale from 1 – 5 please indicate your level of agreement with the following statements. If you have no opinion or the statement does not relate to you, please circle NA.

	Completely Disagree				Completely Agree
I felt welcome when I entered the school.	1	2	3	4	5
The principal at my child's school was an effective leader.	1	2	3	4	5
My child's elementary school was a safe place.	1	2	3	4	5
My child's school was clean and well maintained.	1	2	3	4	5
The teacher helped me to understand what my child was learning.	1	2	3	4	5
The teacher went "above and beyond" in working with my child.	1	2	3	4	5
The teacher had a positive attitude about the school.	1	2	3	4	5
The teacher motivated my child to learn.	1	2	3	4	5
The teacher had high expectations for my child.	1	2	3	4	5
The teacher used multiple approaches to help my child learn.	1	2	3	4	5
The teacher cared about my child.	1	2	3	4	5

About how often in the past year did you:

6a. Have phone contact with your child's teacher?

Daily Weekly Monthly Once per term Less than once per term

6b. Receive email or written communication from the teacher that was specifically about your child?

Daily Weekly Monthly Once per term Less than once per term

6c. Have face-to-face contact with the teacher about your child?

Daily Weekly Monthly Once per term Less than once per term

7a. Were you aware of the Pay for Performance pilot program that was being conducted in your child's school?

YES NO

7b. If yes: What kind of impact do you think it had on your student?

(open text box)

8. Please rank the following items from 1 to 7 in terms of their importance (with 1 being the most important and 7 being the least important) in determining the pay for teachers:

Student academic performance	1	2	3	4	5	6	7
Test scores	1	2	3	4	5	6	7
Teacher's education	1	2	3	4	5	6	7
Teacher's experience	1	2	3	4	5	6	7
Instructional quality	1	2	3	4	5	6	7
Classroom management	1	2	3	4	5	6	7
Parent satisfaction	1	2	3	4	5	6	7

9. Please list any other factors that you feel should be considered in determining the pay for teachers

(open text box)

APPENDIX C: STUDENT OUTCOMES MODEL SPECIFICATIONS AND RESULTS

Model 1: Language Arts

Level-1 Model

2011 LA CRT = B0 + B1*(GRADELEV) + B2*(FEMALE) + B3*(LOWINCOM) + B4*(LIMITEDE) + B5*(SPECIALE) + B6*(WHITE) + B7*(2010 LA CRT) + R

Level-2 Model

B0 = G00 + G01*(P4P) + U0
B1 = G10
B2 = G20
B3 = G30
B4 = G40
B5 = G50
B6 = G60
B7 = G70
B8 = G80

Sigma squared =46.78656; Tau B0=1.13711; Reliability = .796

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	51.564452	0.327015	157.682	574	0.000
P4P, G01	0.432390	0.573526	0.754	574	0.451
For GRADELEV slope, B1					
INTRCPT2, G10	0.391810	0.015508	25.264	160827	0.000
For FEMALE slope, B2					
INTRCPT2, G20	0.334632	0.034481	9.705	160827	0.000
For LOWINCOM slope, B3					
INTRCPT2, G30	-1.049538	0.040497	-25.916	160827	0.000
For LIMITEDE slope, B4					
INTRCPT2, G40	-0.922265	0.074966	-12.303	160827	0.000
For SPECIALE slope, B5					
INTRCPT2, G50	-1.967431	0.053074	-37.070	160827	0.000
For WHITE slope, B6					
INTRCPT2, G60	0.774829	0.055008	14.086	160827	0.000
For LA_10_S slope, B7					
INTRCPT2, G70	0.681448	0.001860	366.417	160827	0.000

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, level-1,	U0	1.06636	1.13711	574	4150.49518	0.000
	R	6.84007	46.78656			

Model 2: Math

Level-1 Model

2011 Math CRT = B0 + B1*(GRADELEV) + B2*(FEMALE) + B3*(LOWINCOM) + B4*(LIMITEDE) + B5*(SPECIALE) + B6*(WHITE) + B7*(2010 Math CRT) + R

Level-2 Model

B0 = G00 + G01*(P4P) + U0
B1 = G10
B2 = G20
B3 = G30
B4 = G40
B5 = G50
B6 = G60
B7 = G70
B8 = G80

Sigma squared =57.13191; Tau B0=2.14215; Reliability = .841

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	52.452156	0.327137	160.337	574	0.000
P4P, G01	0.240963	0.768906	0.313	574	0.754
For GRADELEV slope, B1					
INTRCPT2, G10	-0.106692	0.017144	-6.223	160827	0.000
For FEMALE slope, B2					
INTRCPT2, G20	-0.456425	0.038099	-11.980	160827	0.000
For LOWINCOM slope, B3					
INTRCPT2, G30	-1.314567	0.044810	-29.337	160827	0.000
For LIMITEDE slope, B4					
INTRCPT2, G40	-1.913359	0.082721	-23.130	160827	0.000
For SPECIALE slope, B5					
INTRCPT2, G50	-3.372577	0.058415	-57.735	160827	0.000
For WHITE slope, B6					
INTRCPT2, G60	0.812572	0.061020	13.316	160827	0.000
For MA_10_S slope, B7					
INTRCPT2, G70	0.695976	0.001824	381.628	160827	0.000

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, level-1,	U0	1.46361	2.14215	574	6067.29151	0.000
	R	7.55857	57.13191			