This product is part of the RAND Corporation monograph series. RAND monographs present major research findings that address the challenges facing the public and private sectors. All RAND monographs undergo rigorous peer review to ensure high standards for research quality and objectivity.

# A Big Apple for Educators

## New York City's Experiment with Schoolwide Performance Bonuses

FINAL EVALUATION REPORT

---

Julie A. Marsh, Matthew G. Springer, Daniel F. McCaffrey, Kun Yuan, Scott Epstein, Julia Koppich, Nidhi Kalra, Catherine DiMartino, Art (Xiao) Peng

Prepared for the Fund for Public Schools

RAND EDUCATION

VANDERBILT Peabody College

This research was conducted by RAND Education, a unit of the RAND Corporation, and partners at Vanderbilt University and the National Center on Performance Incentives (NCPI). The research was sponsored by the New York City Fund for Public Schools and NCPI.

**Library of Congress Cataloging-in-Publication Data**

A big apple for educators : New York City's experiment with schoolwide performance bonuses : final evaluation report / Julie A. Marsh ... [et al.].
p. cm.
ISBN 978-0-8330-5251-3 (pbk. : alk. paper)
1. Teachers—Salaries, etc.--New York (State)—New York. 2. Merit pay—New York (State)—New York. 3. School improvement programs—New York (State)—New York. 4. Academic achievement—New York (State)—New York. I. Marsh, Julie A.

LB2842.4.N7B54 2011
371.1009747—dc23

2011023662



*Cover illustration courtesy Getty Images*



Published 2011 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

# Preface

The New York City Schoolwide Performance Bonus Program (SPBP) was a joint program on which both the New York City Department of Education (NYCDOE) and the United Federation of Teachers (UFT) had agreed. Implemented for the first time in the 2007–2008 school year, this voluntary, three-year program provided financial rewards based on school-level performance to educators in high-needs elementary, middle, and high schools. The SPBP represents one of a growing number of pay-for-performance programs being tested around the country.

In 2009, the New York City Fund for Public Schools contracted with the RAND Corporation to independently evaluate the SPBP in partnership with Vanderbilt University and the National Center on Performance Incentives (NCPI). The purpose of this two-year project was to evaluate the program's implementation and effectiveness. This monograph is the final product of this evaluation and should be of interest to policymakers, researchers, and practitioners designing, implementing, or studying pay-for-performance programs or policies.

This research was conducted by RAND Education, a unit of the RAND Corporation, and partners at Vanderbilt and NCPI. Funding to carry out the work was provided by the New York City Fund for Public Schools and NCPI.

# Contents

# Figures

# Tables

# Executive Summary

In the 2007–2008 school year, NYCDOE and UFT implemented the SPBP in New York City public schools. The RAND Corporation, a nonprofit public policy research institution—in a partnership with Vanderbilt University and NCPI, a U.S. Department of Education-funded national research and development center for state and local policy on educator compensation—conducted an independent evaluation of the SPBP for NYCDOE and the UFT, with funding from The Fund for Public Schools and NCPI. This two-year study took place over the period from February 2009 through March 2011 and addressed the following broad research questions:

1. How was the program being implemented from 2007 through 2010?
2. What were the intermediate outcomes of the program?
3. What were the effects of the program on student performance?

This report describes the results of our analyses for all three years of the program, from 2007–2008 through 2009–2010.

## Overview of the Schoolwide Performance Bonus Program

Implemented for the first time in the 2007–2008 school year, the SPBP was a joint NYCDOE-UFT program to explore the utility of an alternative compensation plan as a means of improving student outcomes. The theory of action behind the SPBP was that an incentive pay system

would motivate educators to change their practices to ones better able to improve student achievement. The program's theory of action also hypothesized that the chance to earn a bonus on the basis of school performance could enhance collaboration and that winning bonuses could boost morale—both of which would lead to better outcomes. The theory of action that some leaders articulated held that, in addition to the overall motivational effect of the incentive pay system, rewarding staff differentiated awards on the basis their individual performances could provide extra incentives for change.

SPBP was a voluntary program implemented in high-needs elementary, middle, K–8, and high schools. For schools participating in SPBP, NYCDOE set annual performance targets based on its Progress Reports. The Progress Reports are the district's main accountability tool, measuring student performance and the school environment in all schools in the district. For elementary and middle schools, Progress Report scores are determined in large part by student growth on standardized tests and the school's performance standing relative to other schools, both their peers and all others citywide; scores for high schools are based on graduation and progress to graduation as measured by credit accumulation and performance on Regents Examinations. If a participating school met its annual performance target, as defined by the NYCDOE Progress Reports, the school could receive school-level bonus awards equal to $3,000 per full-time UFT-represented staff member working at the school. The program required each participating school to establish a four-person compensation committee (CC) to determine how the bonus award would be distributed among staff (the CC distribution plan). The committee was made up of the principal, a member designated by the principal, and two UFT-represented staff members elected by staff.

In 2007–2008, Harvard Professor Roland Fryer, with input from the NYCDOE, identified 427 high-needs schools and randomly selected about one-half of these schools for the opportunity to participate in the SPBP in the first year. Fifty-five percent of school staff needed to agree to participate: 205 schools participated in the program in the first year; 198 schools remained in the second year, and 196 in the third and final year.

The program distributed more than $50 million in bonuses over the three years of the study. In year 1 (2007–2008), 62 percent of participating schools won bonuses, resulting in more than $20 million in awards. In year 2 (2008–2009), the numbers increased to 84 percent and more than $30 million. In year 3 (2009–2010), the state raised the thresholds for proficiency, and correspondingly, the percentage of schools earning bonuses decreased significantly. That year, only 13 percent of participating schools earned bonuses, and just $4.2 million were distributed. On announcing the year 3 bonus results in January 2011, NYCDOE simultaneously suspended the program, noting that it would base its decision about whether to resume the program on the results of this study.

## Study Methods

For this evaluation, we collected a variety of qualitative and quantitative data and capitalized on SPBP's experimental design to evaluate program effects. They conducted surveys, site visits, interviews, and case studies to gather data about the implementation.

To understand the implementation and perceived effects of the program from the perspective of staff members closely involved with it, we surveyed all four members of the CCs in the 198 participating schools in spring 2009 (68-percent response rate across all respondents) and 196 participating schools in spring 2010 (72-percent response rate).

To gain an in-depth understanding of the SPBP and staff experiences with the program, we visited seven case-study SPBP schools in spring 2009 and seven additional schools in spring 2010. We selected the case-study schools to represent varying school and student characteristics (school level, student enrollment, borough, student achievement, and student demographics) and school experiences with SPBP (how the school's CC allocated the bonus, whether the school received a bonus, and whether the school voted to participate). Over the two years, we interviewed more than 130 individuals, including principals and assistant principals, teachers, instructional specialists, clinical and student support staff, secretaries, and paraprofessionals.

In addition, to understand SPBP's history, goals, and theory of action, we reviewed official program documents and interviewed NYCDOE and UFT representatives, funders, and other leaders.

To analyze student achievement and compare schools assigned to participate in SPBP with those not assigned to the program, we obtained NYCDOE administrative data for the baseline year (2006–2007) and all three years of implementation. These data included student- and school-level files with biographical, demographic, and student achievement information, along with publicly available data from school Progress Reports. NYCDOE also provided data on school participation, CC distribution plans, and bonus awards for each program year.

Finally, to test whether SPBP affected teachers' attitudes and behaviors (e.g., instruction, collaboration), in spring 2010, we surveyed a sample of teachers in all schools selected to participate in the SPBP and eligible schools that were not selected for participation (57-percent response rate across all respondents).[1]

## Key Findings

### The Study Found No Effects on Student Achievement

**Overall, SPBP Did Not Improve Student Achievement in Any Grade Level.** Analyses of student achievement on the state's accountability tests found that the average mathematics and ELA test scores of students from elementary, middle, and K–8 schools randomly chosen for an invitation to participate in SPBP were lower than those of students from control schools during years 1, 2, and 3. The magnitudes of the estimates, however, are very small and statistically significant only for mathematics in year 3. The results are not significant when we controlled for testing effects from multiple years and subjects. These results were robust under various analytic approaches. Similarly, there were

[1] A PDF file containing the six appendixes to this document is available on the product page for this monograph (http://www.rand.org/pubs/monographs/MG1114.html). These appendixes offer an extensive collection of information on our surveys, bonus distribution, and student achievement data and analysis methods.

no overall effects on Regents Exams scores for high school students in years 1 and 2. We tested for but did not find differential program effects by school size and found no relationship between student achievement and the CC distribution plans for bonus awards among staff.

**SPBP Had No Effects on School Progress Report Scores.** Across all years and all the component scores for the Progress Reports (environment, performance, progress, and additional credit), we found no statistically significant differences between scores of SPBP treatment and control schools and between schools that participated in SPBP each year (regardless of random assignment) and other eligible schools. This lack of effects holds true for elementary, middle, and high schools.

### The Implementation Had Mixed Success in Creating the Optimal Environment for SPBP

SPBP implementation depended on communications about the program to participating schools and staffs, the CC determination of award distributions, the determination of award winners, and the subsequent payout of those awards. Past research on pay-for-performance programs and expectancy theory (Vroom, 1964) suggests that, to achieve the desired program results, these activities must follow a reasonable time line and lead to a high degree of the following key attributes:

- *understanding* of the program, as shown by knowledge of criteria by which incentives are awarded and the amount of money at stake
- *expectancy*, as demonstrated by educators' beliefs that they are capable of doing things that will enable them to achieve the targets
- *valence*, a concept that refers to the belief that incentives are sufficiently valuable or substantial to inspire responses predicted by the theory of action
- *buy-in*, or acceptance of the program and its criteria
- perceived *fairness*.

The following subsections describe the rollout, communication, and the extent to which the implementation achieved the key attributes theoretically necessary for success.

**Basic Procedures Were Enacted as Planned.** The NYCDOE and the UFT generally implemented the SPBP as intended. At the start of each year, schools voted to participate and, later in the year, formed their CCs and submitted bonus distribution plans. Schools had freedom to compensate staff with bonuses as they desired, without interference from either NYCDOE or UFT. Each year, NYCDOE awarded bonuses according to the program guidelines. In the first two program years, district leaders announced the bonus awards within the first few months of the next school year. However, in year 3, announcements were delayed until nearly midway through the subsequent school year.

**Communications Followed the SPBP Design, with Early Problems Being Corrected over Time.** Both NYCDOE and UFT adhered to the communication plan for SPBP, which called for sharing joint NYCDOE-UFT written materials about the program with school staff. According to survey respondents, UFT representatives served as the main conduits for information about SPBP to UFT CC members and teachers. CC respondents reported early misunderstandings but that communications improved over time. Nevertheless, some interviews and teacher survey responses suggest that some misunderstanding of program components remained in year 3, when UFT and DOE reduced the emphasis on communicating about the program.

**The CC Process Was Implemented Fairly and Smoothly, but Some Schools Had Difficulty with the Decisionmaking Process.** Each school formed a four-person CC and generally followed SPBP guidelines on membership and procedures. Most CC members reported that the decisionmaking process was fair, collegial, and inclusive and that achieving consensus was easy; however, some survey and case-study respondents expressed concerns about the process. For example, 44 percent of teachers disagreed or strongly disagreed that teachers' preferences were taken into account when developing the distribution plan. Some CC members whom we interviewed also questioned whether the requirement that UFT-represented staff make up only one-half of the members truly guaranteed an even playing field and spoke

about power differentials that played out between administrators and UFT-represented staff in committee deliberations.

**The Majority of CCs Developed Nearly Egalitarian Award Distribution Plans, Reflecting Strong Preferences Among CC Members for Equal Bonus Shares for All Staff.** Although administrators were significantly more inclined than their UFT counterparts to favor differentiating the bonus awards so that different staff members would be eligible for different amounts, overall there was a strong preference for egalitarian plans among CC members. Further, almost two-thirds of teachers indicated a preference for distributing equal shares of the bonus to all school staff (yet almost the same proportion of teachers also reported that nonteachers should receive a smaller share than teachers). Not surprisingly, the majority of committees developed equal-share distribution plans in both years. The most common individual bonus amount awarded to staff in the plans was $3,000 in all three years. Most staff within a school received the same award amount. In fact, in 2010, 82 percent of staff members were slated to receive the most common or modal award for their school, and even greater equality existed in the early years of the program. Inconsistent with the notion that larger schools might use differentiation in bonus payments to offset "free riders" (i.e., staff who shoulder less than their fair share of the work but still collect a full bonus), award equality increased with school size.

About 31 percent of schools reported using individual performance as at least one of the factors for determining awards. The remaining schools either did not differentiate or reported using only factors related to time or job title but not individual performance. Unequal disbursement at times resulted in resentment within the schools, and some schools with highly differentiated allocation plans one year adopted much more egalitarian plans the subsequent year.

**The Few Schools That Determined Bonus Shares by Individual Performance Tended to Have More Differentiation Among Award Amounts but Did Not Differ from Other Schools in Student Achievement.** Sixty schools reported including individual performance measures, such as staff absences and unsatisfactory staff evaluation ratings ("U-ratings"), as a factor (among many others) for determining individual staff awards. Compared to other schools that did

not report relying on individual performance measures in developing their distribution plans, these schools were more likely to award staff less than the modal awards and more likely to award certain staff members no money. According to the theory of action for SPBP some leaders espoused, greater disparity in awards would help to incentivize performance, have a motivational effect on individual behavior, and increase student achievement. However, even the schools that determined awards by individual performance generally remained cautious about deviating from egalitarian awards and slated 74 percent of staff, on average, for the modal award amount. Moreover, students in these schools—whether elementary, middle, or K–8—did not have higher achievement than those in other SPBP schools.

**A Minority of CC Members and a Few Case-Study Respondents Reported Problems with the Distribution of Bonuses.** The distribution process was reported to run fairly smoothly in the third year of the program, and only a minority of CC members reported problems, such as staff being left off the list of recipients or being upset with a perceived unfair distribution. None of these reported problems reached the level of official appeal in any year of the program. Participants in several case-study schools reported dissatisfaction among staff when decisions and criteria used to differentiate the bonus were viewed as opaque and subjective, or when the schools did not communicate the final distribution plans to staff members.

**The Implementation of SPBP Had Mixed Results in Creating the Circumstances That Foster Success.** We noted the following:

- *Understanding and Awareness.* Staff members reported being aware of the program and generally supportive of it. Most educators surveyed and interviewed greatly appreciated the financial reward and the recognition of their efforts. However, there were persistent misunderstandings about the Progress Reports and other program elements. According to surveys, more than one-third of teachers did not understand key aspects of the program, including the target their school needed to reach, the amount of money their school would receive if they met their target, the source of the funding, and how committees decide on distribu-

tion plans. In the case-study schools, many individuals conveyed misperceptions about various aspects of the programs, and many called for better communication.

- *Expectancy.* In the third year of the program, staff showed some accuracy in their assessment of the efforts required to earn a bonus but, overall, seemed to overestimate the likelihood their school would receive an award. For instance, although many CC members acknowledged needing to improve performance to win a bonus in 2009–2010, a large majority felt certain their school would receive one.
- *Valence.* The majority of teachers and CC members expressed a strong desire to win bonuses and found the financial bonuses motivating, but many winners reported that, after taxes, the bonus seemed insignificant. In fact, almost one-half of the teachers responding to the survey and some staff members in case-study schools indicated that the bonus was not large enough to motivate extra effort. Further, many case-study respondents reported viewing the bonus as a reward for their usual efforts, not as an incentive for changing their behavior.
- *Buy-in.* Buy-in for the program performance measure was limited: The majority of teachers and CC members felt the criteria relied too heavily on test scores.
- *Fairness.* More than one-half of teachers and CC members felt the program was fair to participating schools, and CC members did not report dissatisfaction about unfair distributions. However, some UFT-represented staff expressed dissatisfaction with the composition of the CC committee; more than one-half of teachers and UFT committee members wanted the UFT to have greater than 50-percent representation on the committee. Some staff also felt the Progress Report targets were too high.
- *Time lines.* Three-fourths or more of teachers and CC members suggested that they should have been informed of the distribution plans at the start of the year.

### SPBP Did Not Produce the Intended Effects on Teachers' Reported Attitudes and Behaviors

**SPBP Had no Effects on Teacher Reported Attitudes, Perceptions, and Behaviors.** The survey found no differences in the practices and opinions of teachers in the SPBP group, as compared with those in the control group. On all but one measure related to instructional practice, effort, participation in professional development, mobility, and attitudes, the responses from the two groups of teachers were very similar, and there were no statistically significant differences among them.

**Lack of Results Might Be Due to the Limited Motivational Power of the Bonus.** The theory of action behind the SPBP is that the potential for a bonus would motivate change, but this study did not find such a change. The vast majority of teachers and CC members who received bonuses said that winning the bonus was a nice acknowledgement of their hard work but that it did not influence their performance. In addition, only 39 percent of CC members and 15 percent of teachers reported that not receiving a bonus energized them to improve their practice the subsequent year, and only a very small proportion of both groups actually reported that not receiving the bonus reduced their motivation.

The limited motivational effect of the bonus might have resulted from school staff viewing the award as a reward rather than an incentive: It made them feel appreciated for their hard work, but they claimed that they would have undertaken this work with or without the bonus. The size of the award might also have been a factor. Some stakeholders believed that differentiation among awards could enhance the motivational power of the bonuses. As discussed above, the differentiation especially related to performance was very limited, and it is possible this too contributed to the limited incentivizing effects of the bonus. Finally, other accountability pressures and intrinsic motivation were often perceived to be more salient than the bonus. For example, although 64 percent of teacher survey respondents said the possibility of earning a financial bonus motivated them moderately or to great extent to work hard as a teacher, an even greater percentage reported the same motivational value from the possibility of receiving a high Progress Report grade (without the financial incentive attached)

(75 percent) and achieving or exceeding their school's adequate yearly progress target (77 percent).

**The Lack of Effect Was Also Echoed in Responses About Staff Members' Own Performance.** The majority of teachers, CC members, and case-study respondents said that SPBP did not affect their own job performance or motivation.

**It Is Less Clear Whether Lack of Program Effects on Collaboration or Morale Contributed to the Lack of Intermediate Outcomes.** SPBP's theory of action hypothesized that the chance to earn a bonus on the basis of school performance could enhance collaboration and lead to better outcomes. It also suggests that winning bonuses could boost morale and lead to better outcomes. However, the data on collaboration and morale are mixed. More than 40 percent of participating teachers and CC members reported that teachers' willingness to collaborate and work together, teachers' focus on student learning, and staff morale changed for the better as a result of participation in SPBP. The perception that the program caused the increased willingness to collaborate, however, is not reflected in the comparison with schools not assigned to the program. The reported levels of staff collaboration differed very little between the SPBP and control schools, and the differences that did exist tended to show greater collaboration in the control schools (but the differences were not statistically significant).

## What Explains the Lack of Change Under SPBP?

The findings from the student achievement analysis suggest that, in the three years of implementation, the SPBP did not achieve its ultimate goal of improved student performance. There are several plausible explanations for this result. First, one could argue that the program was too new to produce effects—especially in the first two years. However, if newness were the explanation, one would expect to see at least some positive effects by year 3, but this was not the case.

Second, the findings from the teacher and CC surveys and case-study interviews suggest that several factors important for pay-for-performance programs might not have been in place in all participating

schools. The absence of these factors might have weakened the motivational effects of the bonus. Circumstances that could have weakened SPBP's effects include misunderstandings about the program, uncertainty that teachers could sufficiently change their practices or act in ways to achieve targets, lack of buy-in for the bonus criteria, concerns about timing, and questions about procedural fairness.

A third possible explanation for the lack of positive student achievement effects is that there could be a flaw in the underlying theory of action for SPBP and school-based pay-for-performance programs more generally. As some have argued in the past, motivation alone does not improve schools. Even if the bonus inspired staff to improve practices or work together, it may not have the capacity or resources (e.g., school leadership, social capital, expertise, instructional materials, time) to bring about improvement.

Finally, the lack of observed results could be due to the low motivational value of the Progress Reports and the accountability incentives all schools face. Assessing the true motivational value of the bonus is a difficult task, and our study yielded mixed findings on this topic. While many teachers and other staff reported that the possibility of receiving a bonus did not affect their practices, many nonetheless expressed a desire to earn the bonus and indicated that they were taking it seriously. Nevertheless, many acknowledged that other accountability pressures and incentives (e.g., receiving a high Progress Report grade, achieving adequate yearly progress targets) held the same motivational value as the possibility of receiving a financial bonus.

Hence, while the bonus might have been another factor motivating SPBP staff to work hard or change their practices, they would probably have had similar motivation without it because of the high level of accountability pressure all schools and their staffs face. Consequently, SPBP might not be expected to change behavior or to influence student outcomes.

## Recommendations

Overall, the evidence from this study does not support continuing SPBP. The program did not have the desired effects on student achievement or the necessary intermediate teacher outcomes. There is no consistent evidence that the program motivated educators to change practices or that continuing the program would improve outcomes. Moreover the program was costly, and costs fluctuated dramatically across years.

As was discussed earlier, some implementation problems may have weakened the effects of the bonus and the program overall, and NYCDOE and UFT leaders could consider the following actions to address those:

- Improve communication to participating schools about bonus criteria, how the program operates, potential funds available, and the committee process.
- Adjust the program time line to ensure that all staff know, at the start of the year, the distribution plan and the dollar amounts they can potentially earn.
- Ensure transparency of distribution criteria and plans.
- Provide technical assistance to CC members on the decisionmaking process for determining distribution plans.

However, we do not have evidence that implementing these adjustments will improve the effectiveness of the program significantly. Furthermore, given the consistent failure of recent experiments with similar bonus programs to find positive effects on student outcomes, we suspect a more comprehensive revision of the program and reevaluation of the theory of action is likely to be necessary to achieve the desired outcomes.

Finally, the sharp decrease in the percentage of schools receiving bonuses in year 3 (13 percent) compared to year 2 (84 percent) also suggests a broader lesson for policymakers generally. Although the study did not collect data on responses to the year 3 results, it is hard to imagine how such fluctuations could improve the already weak moti-

vational effect of the program. Prior to implementing a performance-based bonus program, it would behoove leaders to ensure greater stability of the accountability measures on which the bonuses rely.

## Implications for Pay-for-Performance Policies

The results of this evaluation add to a growing body of research from the United States that finds no effects on student achievement of narrow pay-for-performance policies that focus only on financial incentives without other features, such as targeted professional development or revised teacher evaluations. In recent years, a few studies, released well after SPBP started, have demonstrated that several school- and individual-based performance incentive policies had no significant effect on student test results (e.g., Glazerman and Seifullah, 2010; Springer et al., 2009a; Springer et al., 2010a). The NYC SPBP provided one of the few examples of a school-based incentive policy and, like the individual programs, it resulted in few observable effects on students, staff, and schools. The implementation findings, nonetheless, provide insights into the theory of action underlying the policy and the mechanisms expected to yield positive student outcomes—such as improved motivation, enhanced morale, increased collaboration, and improved practice. The study also uncovered much about the process and potential challenges of assigning a small school committee the task of distributing a schoolwide bonus among staff.

Overall, these results yielded several implications relevant to the broader set of pay-for-performance policies that have gained considerable national attention and interest in recent years.

- *Conditions must foster strong motivation.* Our study findings support much of the theoretical literature and other empirical research suggesting that there may be a set of key conditions (e.g., a high degree of understanding, expectancy, valence, buy-in, perceived fairness, and a reasonable time line) needed to bolster the motivational effect of financial incentives. Several of these purported key system components were lacking in SPBP and were

identified by some educators as limiting the ability of program to change their behaviors.

- *It is important to identify the factors that truly affect motivation.* Motivation is the key to the theory of change in pay-for-performance programs. The failure of recent experiments to find effects (Springer et al. 2010a, b) has generally been attributed to the lack of motivational effects of such programs. However, the in-depth data from this evaluation present a somewhat contradictory story. Teachers and other staff reported that the possibility of a bonus was desirable and motivating, but they also reported not changing their behaviors in response to the program. It may be that a desirable award is not enough to actually change behavior. This may be particularly true in the context of high-stakes and high-profile accountability. More teachers reported that receiving a good Progress Report grade was a motivator than did teachers who reported the bonus motivated them. The Progress Report has no direct financial rewards, but it does have high stakes in terms of public opinion and possible negative sanctions. All schools face these accountability pressures, and these pressures may thus reduce the relative value of the additional incentive of the bonus to the point that it does not matter or that it seems merely to substitute for other motivational factors. It may also be that concerns about negative sanctions—again a possibility all schools face—is a greater motivator of true change than a bonus with unknown expected value.[2] Finally, as noted earlier, the limited motivational effects of the bonus reported by teachers and other school staff might not be enough to explain why SPBP failed to change teacher behaviors. Such factors as lack of capacity to change or to identify alternative teaching practices might have prevented change, even if the limited motivation of the SPBP bonus induced teachers to try to improve.
- *Performance-based incentives may face challenges from the micropolitics of school-level implementation.* This evaluation highlighted the

[2] The experience of the POINT experiment in Nashville, Tennessee, is consistent with this conjecture (Springer at al., 2010a).

underlying political tensions inherent in implementing a bonus system. Although many major program elements were implemented smoothly across participating schools (e.g., the formation of CCs), some schools found it difficult to decide how to distribute bonuses among staff (e.g., power differentials played out between administrators and UFT-represented staff during deliberations). The majority of CCs distributed bonuses nearly equally among all staff, and some unequal disbursements exacerbated political tensions within schools. Those seeking to enact similar programs should recognize that the very idea of differentiating pay based on performance will likely challenge deeply ingrained school cultures and norms of collaboration and egalitarianism.

- *Pilot testing and evaluation are essential.* From the outset, NYCDOE and UFT leaders intended to implement the SPBP on a pilot basis. Resource constraints prevented the planned scale-up within the initial group of schools identified as eligible for this pilot program, providing evaluators with three years to compare outcomes and practices across a set of participating and comparable high-needs control schools. Implementing the program on a small scale and including randomized treatment and control groups for three years (perhaps not intended at the outset) provided valuable information to inform future decisions about an essentially untested policy innovation. Leaders created the opportunity to gather comprehensive data on the implementation and outcomes to assess the merits of continuing and further scaling up this policy. The pilot period and evaluation also provided leaders an opportunity to reexamine the theory of action and assumptions underlying the SPBP.

# Acknowledgments

Many individuals contributed to this monograph. We thank all the staff at the NYCDOE and UFT for their support throughout this evaluation, particularly Terry Bowman, Joanna Cannon, Vicki Bernstein, Oliver Ho, Tolani Adeboye, Dominique West, Kimberly Howard, Martin Kurzweil, Mike Manner, and George Stydahar at NYCDOE and Michael Mendel and Susan Amlung at UFT. We are also grateful to the sponsors of this research, the Fund for Public Schools and the National Center on Performance Incentives, which is funded by the U.S. Department of Education's Institute of Education Sciences. We are indebted to the many administrators, principals, teachers, paraprofessionals, and other school staff who participated in the study and shared their valuable time and insights with us. We also thank the principals and teachers who pilot tested our surveys.

The project would not have been completed without the assistance of many colleagues from RAND, including Melissa Bradley and her team within the Survey Research Group; Bas Weerman, Tanya Gutsche, and Emily Payne from Multimode Interviewing Capability; Sharon Koga, Chris Dirks, Yvonne Torres, Elizabeth Steiner, Al Crego, Brian Stecher, Susan Bodilly, Robert Hickam, Louis Ramirez, Sheila Logue, and Lance Tan, and from Vanderbilt University's Peabody College, including Ryan Balch, Cate Gardner, Jessica Lewis, Amie Rappaport, and Jeffrey Springer. We thank Laura Hamilton, Lou Mariano, Sean Reardon, and Alix Gallagher for their thoughtful reviews and comments on this report and earlier versions of this monograph and Phyllis Gilmore for her edit of the final version.

## Abbreviations

| | |
|---|---|
| AYP | Adequate Yearly Progress |
| CC | compensation committee |
| CSA | Council of School Supervisors and Administrators |
| ELA | English language arts |
| ELL | English-language learner |
| GEEG | Texas Governors' Educator Excellence Award Grants |
| ITT | intent-to-treat |
| K–8 | kindergarten through 8th grade |
| MOU | memorandum of understanding |
| NCLB | No Child Left Behind |
| NCPI | National Center on Performance Incentives |
| NYCDOE | New York City Department of Education |
| POINT | Project on Incentives in Teaching |
| SPBP | Schoolwide Performance Bonus Program |
| TAP | Teacher Advancement Program |
| TEEG | Texas Educator Excellence Award Grant |
| UFT | United Federation of Teachers |
| USDOE | U.S. Department of Education |

CHAPTER ONE

# Introduction

In the 2007–2008 school year, the New York City Department of Education (NYCDOE) and the United Federation of Teachers (UFT) implemented a pay-for-performance program called the Schoolwide Performance Bonus Program (SPBP). In accordance with the memorandum of understanding (MOU) that established the SPBP and called for an independent evaluation, NYCDOE and UFT contracted with the RAND Corporation (in partnership with the National Center on Performance Incentives [NCPI] and Vanderbilt University) to evaluate the implementation and effects of this program. The evaluation study was funded by the Fund for Public Schools and NCPI, which is funded by the U.S. Department of Education's Institute of Education Sciences. The two-year study started in February 2009 and was designed to examine (1) program implementation, factors affecting implementation, and implementation progress over time; (2) how SPBP affected school, staff, and student outcomes; and (3) the links between implementation and desired outcomes.

This monograph describes the data collection and analysis conducted during the two-year evaluation and provides findings on the three years of the SPBP.

## A Brief Overview of the Schoolwide Performance Bonus Program

Implemented for the first time in the 2007–2008 school year, SPBP was a joint program of NYCDOE and UFT. Established as a two-year

pilot for the 2007–2008 and 2008–2009 school years, the program was extended in 2009 for a third year (2009–2010) as a result of additional negotiations between NYCDOE and UFT. The program was suspended in January 2011 on announcement of payment of year 3's bonuses.

The voluntary program provided financial rewards to educators in high-needs elementary, middle, kindergarten through 8th grade (K–8), and high schools.[1] Each school needed the approval of 55 percent of its UFT-represented staff members to participate in the program each year. A participating school could receive a school-level bonus equal to $3,000 multiplied by the total number of full-time UFT-represented staff members working at the school,[2] if the school met its annual performance target. That target was defined by the NYCDOE accountability program and was determined in large part by student growth on standardized tests. The program required each participating school to establish a four-person compensation committee (CC) to determine how to distribute the bonus among staff members. The committee was made up of the principal, a member designated by the principal, and two UFT-represented staff members elected by staff.

In 2007–2008, 427 high-needs schools were identified for the program; about one-half of these were randomly selected to be offered the opportunity to participate in the SPBP. Over time, a few schools that accepted this opportunity dropped out or closed, but most of the schools that agreed to participate did so for all three years. Ultimately,

---

[1] For high schools, high need was measured using the average proficiency ratings of actively enrolled students on the 8th grade New York State English language arts (ELA) and mathematics exams. For middle schools, it was measured using average proficiency ratings on the 4th grade ELA and mathematics tests. For elementary schools, it was calculated on the basis of poverty rates, other student demographic characteristics, and the percentages of ELL and special education students in the schools.

[2] UFT-represented refers to employees covered by one of several NYCDOE-UFT negotiated contracts. These include teachers, social workers, counselors, school nurses, psychologists, school security personnel, teachers' aides, nonsupervisory education personnel, adult-education employees, speech therapists, school secretaries, and substitute teachers. While it does include paraprofessionals working side by side with teachers, it does not include school aides, who are covered by employee union DC37.

205 schools participated in year 1 (2007–2008), 198 schools in year 2 (2008–2009), and 196 schools in year 3 (2009–2010).

## Purpose of the Evaluation

The purpose of our two-year project was to independently evaluate SPBP. Specifically, the evaluation was designed to answer the following questions:

1. How was the program implemented?
2. What were the intermediate outcomes of the program?
3. How did the program affect student performance?

## Data Collection and Analysis

As described in more detail in Chapter Three, we collected and analyzed a variety of qualitative and quantitative data over the course of the evaluation, including interviews with representatives from NYCDOE, UFT, funders, and other leaders (years 1 and 2 of the evaluation); site visits to 14 SPBP schools (seven each year); surveys of CC members in all SPBP schools (years 1 and 2); surveys of a sample of teachers in all SPBP and in eligible schools that were not selected for participation (year 2); documents from NYCDOE and UFT; administrative data from the 2007–2008, 2008–2009, and 2009–2010 school years (participation, distribution plans, bonuses); and student achievement data from SPBP schools and from eligible schools that were not selected for participation.

## Organization of the Report

In the remainder of this monograph, we first examine pay-for-performance programs more broadly and then delve into the design, implementation, and effects of the New York City SPBP. Chapter Two, provides background on pay-for-performance programs; a summary of past

research on programs similar to New York City's SPBP; and details on the history, design, and theory of action underlying the SPBP. Chapter Three describes our research methods, including the research questions, conceptual framework, and data collection sources and analyses. Chapters Four through Six provide detailed findings on the three years of program implementation, including information about general attitudes, communication, the committee process, bonus distribution, responses to bonuses, and perceptions about effects. Chapter Seven examines the effects of SPBP on school-level Progress Report scores and on student test results in SPBP and control schools. Chapter Eight compares teacher reports on attitudes and classroom practices in SPBP and control schools. The document concludes in Chapter Nine, with a summary of findings, recommendations, and implications.

Finally, a PDF file containing the six appendixes to this document is available on the product page for this monograph (http://www.rand.org/pubs/monographs/MG1114.html). These appendixes offer an extensive collection of information on our surveys, bonus distribution, and student achievement data and analysis methods.

CHAPTER TWO

# Background on Pay-for-Performance Programs and the New York City SPBP

This chapter provides background on performance incentive programs and past research, followed by detailed information about the New York City SPBP.

## Background on Pay-for-Performance Programs

Pay-for-performance programs and policies tie employee compensation to performance on the job and have long been common in the commercial sector as a means of providing incentives and rewards to meet specific production targets or goals. In recent years, pay-for-performance has also received substantial attention as an option for meeting goals in the public sector.

Although most school districts continue to tie educator pay to years of service and education level, many states, districts, and schools in recent years have experimented with alternative compensation systems that include not only performance-based pay but also bonus pay for acquiring new knowledge and skills, teaching particular subject areas, and working in hard-to-staff schools (Johnson and Papay, 2009; Podgursky and Springer, 2007; Springer, 2009). Proponents of performance-based compensation argue that the current single salary scale provides weak incentives for educators to act in the best interests of students and to work to improve teaching practices. Advocates see systems linking pay to performance as a powerful way to reward and motivate

educators to work and improve instruction, and a way to attract and retain high-quality teachers (and to encourage low-quality teachers to exit) (e.g., Solmon, 2006; Stronge, Gareis, and Little, 2006).

Opponents, on the other hand, contend that performance-based compensation is not appropriate for education. They observe that education is a highly personalized, complex, and context-dependent endeavor and that isolating contributions to learning is difficult, that multiple individuals contribute to outcomes, and that output is difficult to measure (e.g., Murnane and Cohen, 1986). Critics further argue that extrinsic incentives may conflict with or undermine the intrinsic motivations of educators (e.g., Deci, 1971; Fehr and Gatcher, 2001) and lead to potentially negative effects, such as weakening collaborative school environments (e.g., Pechthalt, 2007). Finally, others assert that increasing motivation will not fundamentally improve education; rather, they argue, the true need is to enhance instructional capacity (e.g., Elmore, 2004; McLaughlin, 1987).

### Historical Antecedents in Education

Interest in performance-based compensation in education has a long history in the United States. The current wave of reform efforts follows earlier experimentation in the late 19th and early 20th century, as part of Fredrick Taylor's scientific management movement; in the 1960s, sparked by the Russian launch of Sputnik and concerns about the quality of American education; and in the 1980s, following the publication of *A Nation at Risk* (National Commission on Excellence in Education, 1983; Johnson and Papay, 2009; Springer, 2009). In the 1980s, a number of school districts around the country attempted to modify their teacher salary schedules by introducing "merit pay." At that time, merit pay took the form of awarding financial bonuses to teachers whom principals identified as high-performing based mainly on their classroom observations of the teachers. However, these observations were rarely based on common standards or rubrics, and inter-rater reliability was weak at best. Moreover, these pay systems were inadequately funded, and it was not unusual for the money to run out too soon. As a result, some teachers whose ratings warranted bonuses received none (Koppich and Rigby, 2009).

### Current Pay-for-Performance Programs

Although controversial, current programs have gained considerable political and financial support, including endorsements from President Barrack Obama and Secretary of Education Arne Duncan, and increased federal funding. For example, the Teacher Incentive Fund, started in 2006 under President George W. Bush, supports performance-based teacher and principal compensation systems in high-needs schools around the country and has grown substantially (from $99 million in 2006 to $439 million in 2010). The Obama administration's Race to the Top initiative also supports the spread of this reform by requiring, as one of its "state reform conditions criteria," that states and their districts use rigorous evaluations to "inform decisions regarding compensating, promoting, and retaining teachers and principals, including by providing opportunities for highly effective teachers and principals . . . to obtain additional compensation and be given additional responsibilities" (USDOE, 2009, p. 9).

Research indicates that fewer than 6 percent of public schools tied pay to performance in the 1999–2000 academic year, but by 2003–2004, approximately 13 percent of school districts had some form of an incentive program (Podgursky and Springer, 2007; Taylor, Springer, and Ehlert, 2009). Like New York City, several large districts (e.g., Denver, Dallas, and Houston) and states have implemented teacher pay-for-performance systems. For example, as of 2009, Florida, Minnesota, and Texas allocated more than $550 million to such programs (Springer and Winters, 2009). In 1999, the Denver Public School District launched a pilot program in 16 schools that linked teacher pay to the achievement of measurable student performance objectives. The program was subsequently adapted to include multiple ways for teachers to earn bonuses—including increasing their knowledge and skills, teaching in hard-to-staff schools, receiving good performance evaluations, and improving state test scores—and was scaled up to all schools in the district.

In 2006, Texas Governor Rick Perry led the adoption of a pilot pay-for-performance initiative, known as the Texas Governors' Educator Excellence Award Grants (GEEG), which was subsequently scaled up as the Texas Educator Excellence Award Grant (TEEG) program.

Both the GEEG and TEEG programs consisted of several initiatives, including a school-based award program and individual teacher awards (Springer et al., 2009). In 2008, the state of Texas implemented a third performance-pay program known as the District Awards for Teacher Excellence program, which allocates approximately $197 million in state funds each school year to districts for the implementation of locally designed pay-for-performance plans.

Another significant program operated throughout the country is the Teacher Advancement Program (TAP) developed by the Milken Family Foundation in 1999. TAP is a comprehensive school reform model which, among other things, financially rewards teachers based on student performance, increased roles and responsibilities, and classroom teaching performance. Springer, Ballou, and Peng (2008) notes that, at the time of that report, TAP operated in over 125 schools in 50 districts in 9 states, with at least 10 additional states seeking to implement the program.[1] In the 2010–2011 school year, TAP leaders reported operating in schools in 66 districts in 11 states, affecting more than 100,000 students and 10,000 teachers (National Institute for Excellence in Teaching, 2011).

### Theoretical Basis for Pay-for-Performance Programs

Many pay-for-performance programs are grounded in expectancy theory. Put forward by Vroom (1964), this theory suggests that individuals are motivated to engage in certain behaviors when they believe that those behaviors contribute to an organization's goals (expectancy), when the individuals further believe the behaviors will be rewarded (instrumentality), and when the individuals value the rewards (valence). In the pay-for-performance context, expectancy is thus the extent to which educators believe they can improve student performance to meet program goals; instrumentality is the extent to which educators per-

[1] For more details about these and other programs, see Podgursky and Springer (2007). While we focused our literature review on pay-for-performance programs for educators, and school-based bonus programs in particular, financial incentives in education are not limited to educators. Recent (and highly controversial) programs have been implemented that pay students (Fryer, 2010) and parents (Riccio et al., 2010) for performance, attendance, and a variety of other behaviors and outcomes.

ceive that meeting performance goals will result in receiving a bonus; and valence is the degree to which educators value the performance payments (Milanowski, 2000). This theory suggests that a pay-for-performance program that has a high degree of expectancy, instrumentality, and valence may be a strong motivational and reform tool.

### Variations in Pay-for-Performance Program Design

Pay-for-performance programs for educators vary widely in their design and structure. Here we draw primarily on the framework defined by Springer and Balch (2009) to describe the dimensions along which these programs vary.

First, programs may vary in terms of the unit of accountability, the entity responsible for meeting program goals and eligible for program rewards. An individual unit of accountability targets the educator, while a group unit of accountability targets some group of educators, such as teachers of the same grade, department, or any other subschool unit or of the school as a whole. Finally, hybrid models of accountability have some elements of group and individual accountability. That is, part of the reward comes from individual teacher performance and behavior, while part of the reward comes from the group.

A second source of variation is the incentive structure or how goal attainment is defined and how awards are allocated. There are two main types of incentive structures. Rank-order tournaments limit awards to a fixed percentage of the eligible participants (e.g., the top 10 percent of schools receive a bonus). In this way, it is the relative performance of individuals or groups that matters, rather than the absolute performance. In fixed-performance systems, a standard threshold is set, and any participant (e.g., a teacher or school) meeting that threshold receives a reward.

Third, programs can vary in terms of what they measure and on what basis bonuses are awarded. Some programs measure educational inputs or behaviors and activities believed to improve teaching and learning. Commonly used as the basis of performance-pay programs developed prior to 2002, these inputs can include attending professional development, taking on special responsibilities, and taking less

leave. Other programs measure educational outputs or the effects of educational inputs, such as improved test scores.

Fourth, programs may use different performance standards and thresholds. Linear models are continuous and give rewards proportional to the degree of gain in a particular outcome, such as the change in average standardized scores. Step functions give rewards only when gains meet certain thresholds. There may be one or many such thresholds. For example, if schools are "graded," there may be a reward for improvements from a "C" to a "B" (defined by particular thresholds) but no reward from a low "B" to a high "B." Limited linear models are hybrids of these approaches.

Fifth, programs may vary in terms of the size of the incentive, the distribution method, and payout frequencies. Program designers must balance the need to offer a reward that is large enough to motivate effort but not too large as to make the program financially infeasible. Incentives in the U.S. range from a minimum of $250 to a maximum of $12,000, although most are under $3,000 (Taylor, Springer, and Ehlert, 2009). The distribution method refers to how eligibility for an award and the size of the award are established for a group unit of accountability. An egalitarian approach is widely inclusive in determining eligibility and distributes the award equally among eligible participants. Nonegalitarian approaches limit eligibility and make unequal awards, typically allocating more to the top performers and less or none to lower performers. Payout frequencies refer to how often awards are made and when awards are made relative to the time that performance was assessed. We return to this typology later in describing SPBP.

## Research Findings About School-Based Bonus Programs

Under the New York City SPBP, the school was the unit of accountability, and the schools determined their own distribution methods. We next turn to recent literature on similar programs to examine (1) how educators distribute bonuses among staff, (2) educators' attitudes toward these programs, and (3) how monetary rewards and the pro-

grams as a whole affect educators and students.[2] It is worth noting that, at the time New York City introduced SPBP, the pay-for-performance literature was rapidly evolving, and many of the more-recent studies documenting effects were not yet available.

### Trends in Teacher-Developed Allocation Plans

The GEEG program in Texas is one of very few bonus programs that, like New York City's SPBP, allows school staff to determine how the school bonuses should be distributed. Taylor, Springer, and Ehlert (2009) reports that teachers in general designed egalitarian plans for their schools (i.e., the range of bonus amounts in the schools was relatively small, compared to what was allowable under program guidelines). Moreover, schools with more-experienced teachers were more likely than schools with less-experienced staff to have egalitarian plans. Similarly, schools whose teachers' salaries were similar were more likely than schools with uneven salary distributions to have egalitarian plans. According to the authors, "this observation suggests a possible policy tension between incentives that are strong enough to elicit a behavioral response from teachers and the need for teacher acceptance and participation in such plans" (p. 218).

---

[2] We did not review earlier literature on merit pay experiments or surveys of teacher attitudes toward incentive pay from the 1980s. For further details on these programs, see for example, Ballou and Podgursky, 1993; Murnane and Cohen, 1986; Johnson, 1986; Odden and Kelley, 2002; Podgursky and Springer, 2007. As noted earlier, these studies generally found that the programs often involved small bonuses, were not always well understood by participants, and based decisions on less-than-thorough evaluations and measures. Often, teachers did not know the basis for bonuses, and as a result, the researchers involved did not consider these programs to have provided sufficient incentives to change teacher practices. Also, our review focused primarily on school-based programs in the United States. For information on studies of individual performance-pay programs and international programs, see Podgursky and Springer, 2007, and the recent report on The Project on Incentives in Teaching (POINT) experiment in Nashville Public Schools (Springer, Ballou, Hamilton, Le, Lockwood, McCaffrey, Pepper, and Stecher, 2010a). In fact, other than POINT, some of the most rigorous outcome studies come from international programs in Kenya, India, Israel, and Mexico; however, these often employ incentive structures that are very different from those in the United States (e.g., much larger monetary bonuses) and also take place in settings that are quite different from urban districts in the United States.

Taylor and Springer (2009) uses data from the GEEG program to explore incentive design not only from the perspective of the employer—by examining changes in student test scores and teacher retention—but also from the perspective of the employee—by examining the preferences revealed in the incentives teachers design for themselves. The authors found that, when given the opportunity, teachers design relatively egalitarian, group-oriented incentive pay plans. In turn, these incentives did not appear to have induced any significant changes in student test scores, although they did have significant effects on teacher turnover.

### Teacher Attitudes and Effects on Motivation

Research indicates that educators' experience and age affect attitudes toward performance-based financial rewards. In Texas, for example, inexperienced teachers conveyed more support for the rewards than did more experienced teachers (Springer et al., 2009a). Coggshall et al. (2009) reports that younger, "Gen Y" teachers tend to believe that teachers should be monetarily rewarded for success and are more in favor of rewarding performance with bonus or incentive payments than older teachers. However, both Gen Y and older teachers are more likely to support an incentive program that assesses awards at the school level, than one at the individual level.

Several studies examine the motivational effects of school-based bonus programs and report similar findings: Monetary bonuses can motivate teachers but are only one of several influences on teacher attitudes and behavior. Relying on large-scale surveys of teachers in school-based bonus programs in Kentucky and North Carolina, Heneman and Milanowski (1999) reports that teachers appeared to perceive these programs as opportunities to achieve personal, or intrinsic rewards, in addition to financial rewards. Intrinsic rewards included personal satisfaction, the satisfaction of seeing students improve, and positive recognition from achieving the goals. Simultaneously, teachers reported being motivated by the fear of sanctions, such as negative publicity or loss of control over the school. Kelley (1999) examines the motivational effects of several bonus programs in Kentucky, North Carolina, Colorado, and Maryland and reported similar findings. According to

Heneman and Milanowski (1999), the evidence "suggests that the use of bonuses, in conjunction with these other outcomes, has strong motivational potential for changing teachers' behavior in the desired direction of improving student achievement" (p. 338).

Nevertheless, the research is not overwhelmingly positive about the effects on motivation, and Henemen and Milanowski's conclusions should be interpreted with caution. For example, in one substudy of 16 schools participating in a schoolwide bonus program in Kentucky, teachers reported that, although the bonus was a nice acknowledgment of their work, they were not primarily motivated to change their teaching practices because of the potential to receive it (Kelley, 1998). Further, some have criticized the studies in Kentucky and North Carolina for basing their optimistic conclusions about the motivational effects of school-based performance rewards on self-reported data and data collected "as an after-the-fact response to a new program," instead of on an assessment comparing motivational levels before and after the intervention over the long term (Malen, 1999, p. 388).[3]

Research also indicates that several factors mediate the motivational effect of bonus programs and, in some cases, cause teachers to withdraw their support. First, some bonus programs have been associated with higher teacher stress. In the study of Kentucky and North Carolina programs, Kelley (1999) reports that 87 percent of Kentucky teachers and 72 percent of North Carolina teachers reported feeling more job pressure and stress as a result of participating in the bonus programs. These undesirable outcomes may cause teachers to question whether the risks and negative consequences are worth the potential benefits (Heneman and Milanowski, 1999).

Second, some research shows that teachers withdraw support if they feel that the bonus program is not being fairly administered, either because they perceive that the goals are not appropriate or that bonuses are not allocated or awarded fairly (Heneman and Milanowski, 1999). Consistent with expectancy theory, Kelly and Finnegan (2003) reports that perceived fairness of the bonus program was the largest predictor of teacher expectancy or the belief that individual effort would result

[3] The latter is also referred to as a *pre-post assessment*.

in meeting goals. Researchers also specified two types of fairness that mattered for expectancy: substantive fairness (program design accounts for differences in student population and school resources) and procedural fairness (developing, communicating, and adhering to rules so that participants know what is expected of them) (Kelley, Odden, Milanowski, and Heneman, 2000).

Finally, the size of the incentive or bonus may affect the motivational value. While there is no definitive evidence on the optimal size of a bonus, some research suggests that small awards may compromise the motivational value of the bonuses (e.g., Kelley et al., 2000; Malen, 1999; Chamberlin et al., 2002; Heinrich, 2007). According to Odden (2001), research in the private sector has found that annual bonuses need to be at least 5 to 8 percent of salary to influence a worker's motivation.

### Effects on Staff Collaboration and Practice

Research is inconclusive about the effects of pay-for-performance programs on staff collaboration. Several studies of an incentive program implemented at one California charter school provide evidence on this topic. Milanowski and Gallagher (2000) found that teacher perceptions of collegiality decreased initially with the implementation of the individual skill-based evaluation and compensation plan, but perceptions of collegiality and community returned over time. However, the latter observation could have resulted from the school's subsequent adoption of a school-based incentive program for the entire staff (Kellor, 2003).

Teacher involvement in the design of the pay-for-performance system has also been found to influence perceptions of collegiality or competition, but the limited evidence is conflicting. Kellor (2003) points out that, in a program mostly designed by teaching staff, decisions about individual pay appeared to introduce friction in a usually collegial community. However, in an examination of a teacher-developed pay-for-performance program in Texas, Springer et al. (2009a) reports that teaching staff did not believe the incentive program reduced collaboration or collegiality. This view was held by both those who received bonuses and those who did not.

Although little empirical literature delves into the ways performance incentives influence teacher behavior in the classroom, studies of performance-based accountability policies—which are a component of output-based pay-for-performance programs—provide ample evidence that these policies affect classroom behavior in desirable and not-so-desirable ways. Teachers not only tend to spend more time teaching the subjects and specific instructional content that are tested but also tend to focus more on standards and achievement, use testing information and data more extensively to inform instructional changes, and endeavor to improve their own teaching practices in the tested subjects (Hamilton et al., 2007; Hannaway and Hamilton, 2008; Stecher et al., 2009). However, these beneficial behavioral changes are often accompanied by a proportional decrease in instructional time and emphasis on untested subjects (Hamilton, 2004).

Select studies also find that teachers often respond to high-stakes accountability policies by spending more time teaching test-taking strategies and providing differential treatment to students in ways designed to increase overall test scores and not necessarily learning for all students (e.g., targeting "bubble kids" or students who are likely to "pass" or score as proficient) (Hamilton, 2004; Booher-Jennings, 2005; Reback, 2008; Neal and Schanzenbach, 2010).[4] Other studies have similarly documented behaviors intended to "game" the system, such as systematically reclassifying or excluding from testing students who are likely to perform poorly, altering student test data, or reclassifying students to particular subgroups in ways that alter either their eligibility or how their results are factored into accountability measures (e.g., Jacob and Levitt, 2003; Koretz et al., 1996; Figlio and Getzler, 2002).

Some of this research is consistent with the Kelley (1999) study of schoolwide bonus programs in Kentucky and North Carolina in which large majorities of teachers reported that, to achieve school accountability goals, they spent more hours teaching, changed their teaching content, and worked collaboratively with their colleagues to plan

[4] Other studies do not find strong evidence in support of the bubble-student or educational triage hypothesis. See, for example, Springer (2008), Ballou and Springer (2009), and Ladd and Lauen (2010).

and implement teaching changes. Nevertheless, it is unclear whether these reported changes are attributable to the schoolwide bonus or the broader standards-based reforms occurring in these places.[5]

### Effects on Staff Mobility

Few studies empirically examine how bonus programs affect teacher turnover. One such study of the Texas GEEG found no evidence that schools in the program experienced systematic teacher turnover (Taylor and Springer, 2009). However, there were significant differences in the first year of the program for experienced teachers, in the second year of the program for beginning teachers, and in the third year for teachers as a whole. However, there were significant differences among experienced teachers for year 1, among beginning teachers for year 2, and among teachers as a whole for year 3. In these three cases, teachers who did not receive a bonus were significantly more likely to leave their schools under all three incentive structures, but the effect was much less pronounced in schools that designed plans with only teacher incentives than it was in schools with at least some group incentives. Additionally, in all three years of the GEEG program, when the distribution plan was designed to reward all teachers equally, the failure to receive a share was a strong predictor of teacher turnover.

A recent study of the TAP program in Chicago, which provides individual performance-based bonuses to teachers, found that the program had no discernable influence on the rates of teacher retention during its second year of implementation (Glazerman and Seifullah, 2010).

### Effects on Student Achievement

The effects of school-based pay-for-performance programs on student achievement are generally mixed, particularly in the United States. Pod-

[5] A random-assignment study of a schoolwide pay-for-performance program in rural Kenyan elementary schools in which bonuses were quite substantial (21 to 43 percent of monthly salary) found that participating teachers provided more after-school test preparation but that their homework assignments and classroom pedagogy did not differ from those of control teachers (Glewwe, Ilias,-[ and Kremer, 2004). A recent study of Nashville's POINT found that individual incentives had no effect on what teachers reported doing in their classrooms (Springer et al., 2010a).

gursky and Springer (2007) reviewed a variety of domestic and international incentive programs and the studies and reports that associated several of these programs with positive student outcomes. For instance, the Dallas Independent School District implemented a districtwide accountability program that gave all teachers a small monetary reward tied to schoolwide achievement to improve the performance of all the district's schools. Analysis showed that student achievement in Dallas increased relative to other districts in Texas. Nevertheless, given that the intervention did not allow for more rigorous research designs, and many other facets of the district were changing at the same time, these findings are not very robust or informative.[6]

Another study (Figlio and Kenny, 2007) merged national data from several surveys on school staffing policies, student achievement, and other variables to assess the variation and effects of incentive pay in schools. Podgursky and Springer paraphrased the conclusions of the earlier study:

> The effects of even modest doses of incentive pay were statistically significant in both public and private schools. . . . The effect of a high level of implementation of incentives relative to none at all impacts achievement comparable to a one standard deviation increase in days absent for the average student and an increase in maternal education of three years. (Podgursky and Springer, 2007, p. 31)

On the other hand, a number of recent studies have found no evidence of systematic links between performance pay and student

[6] Additional studies, not reviewed here, examine the effects of individual-based performance bonus programs. For example, a recent study of the second year of the TAP program in Chicago found it did not affect student test scores (Glazerman and Seifullah, 2010). Similarly, the study of POINT in Nashville Public Schools found no effect of individual incentives had no effect on test scores overall (Springer et al., 2010). Also, a meta-analysis of 45 studies of incentive programs within private and public businesses, as well as colleges and universities (not K–12 schools), found that group-directed incentives had significantly greater influence on performance (twice the effect) than did individually directed incentives (Condly, Clark, and Stolovitch, 2003). Authors caution, however, that the unequal sample sizes for the studies looking at group-based incentives (n=9) versus individual incentives (n=55) may have affected the statistical results.

achievement, as measured by test scores. For example, analysis of student achievement levels and student achievement growth data in the third year of the TEEG program found no evidence that it had a systematic treatment effect on student achievement (Springer et al., 2009a). Furthermore, neither the design features of the program nor size of teacher bonuses moderated the effects (Springer et al., 2009a). Analyses undertaken as part of the study of the Texas GEEG program produced similar results (Springer et al., 2009b).

When there were no effects on achievement, Podgursky and Springer (2007) found that teachers did in fact respond to the incentives in place but that these incentives did not tie rewards to achievement goals and were therefore not well designed. For example, in a bonus program implemented in an alternative high school in Michigan, bonuses were tied to increased course completion rates rather than to achievement in the courses. In comparison to another alternative high school without such a program, this school's course completion rate did increase—indicating that teachers responded to the incentive—but the school's overall student pass rates and grade point averages dropped as lower-achieving students were induced to stay in school.[7]

These results did not hold in the recent randomized experiment on incentives for individual teachers. One study evaluated the TAP school program and found that, in the context of the participating schools, the prospective of individual awards did not induce changes in teacher behavior (Glazerman and Seifullah, 2010). Similarly, in POINT, middle school mathematics teachers randomly assigned to the invention could earn bonuses of up $15,000, but these incentives did not elicit behavioral differences from teachers in the control group.

As our work was nearing completion, two other studies on the effects of SPBP were released. The first, which appeared in *Education Next*, tested the program's effects on student achievement in years 1 and 2 (Goodman and Turner, 2011). The study used school-level data weighted by the number of tested students and found that the program had no overall effect on student achievement.[8] The study found that

[7] See Podgursky and Springer, 2007, for more-detailed reviews of these and other studies.

[8] The findings we report in Chapter Seven, which use student-level data, are similar.

the program improved student achievement in small schools but that this effect was significant only for mathematics in year 1. We could not replicate this effect using student-level, Progress Report, or teacher outcome data.[9]

The second report appeared in the National Bureau of Economic Research Working paper series. The authors tested how SPBP affected student achievement, attendance, behavioral incidents, course grades, and graduation rates (Fryer, 2011). It also tested how the program affected teacher retention in the district, personal absences, and responses to the district-administered school survey, which measures school environment and was one component of the performance measures used in determining bonuses. The study used student-, teacher-, and school-level data from all three years of SPBP and found no evidence that the program increased student achievement. The study found statistically significant negative results for middle school students on both the mathematics and ELA tests and significant negative results for pooled elementary and middle school students on the mathematics tests. The study found no significant effects for high school Regents Examinations, although all estimated effects were negative. It did, however, find negative effects for graduation rates. It also found significant negative effects for large elementary and middle schools in achievement but, unlike Goodman and Turner (2011), did not find positive effects for small schools. The study found that the program did not affect attendance, student grade point averages, or behavioral problems, other than a significant increase in behavioral problems in SPBP middle schools. It also found no effects on any teacher outcomes.

### Enabling Conditions

Although not many studies focus on school-level conditions and efforts associated with winning a bonus, at least one qualitative study of the Kentucky school-based performance award program identified several enablers of success (Kelley, 1998). Most notably, schools that won the

[9] Contributors to the differences between our findings might be that we lacked the authors' data on school sizes and did not know which school they included in their treatment and control groups.

bonus were more likely to engage in particular strategies, such as aligning the curriculum to state test and curriculum guides, incorporating test-taking strategies into the curriculum, and providing teachers with professional development opportunities focused on knowledge and skill deficits. These schools were also characterized by well-informed teachers with strong professional ties and knowledge about how to use information from the state to guide improvement.

Kelley (1998) further reported that strong principal leadership also played an important role in motivating schools to improve and "substituted for teacher leadership in the schools with slightly lower levels of teacher knowledge, skills, and professional savvy" (p. 320). The level of schooling also appeared to affect outcomes: elementary schools were more likely to win rewards because of what respondents described as "smaller organizational structure, cross-disciplinary nature, and greater malleability and willingness on the part of teachers to try new things" (p. 316).

#### Summary

Past research has found mixed evidence of the motivational effects of school-based bonus programs and has indicated that motivation is often mediated by perceptions of fairness and the size of the bonus. Research is also inconclusive about the effects performance-based incentives have on staff collaboration, while some of the broader accountability literature suggests that there are some potential desirable and undesirable effects on classroom practices. There is also limited and mixed research evidence on how these programs affect student achievement. These findings are important to keep in mind when considering the design, implementation, and outcomes of the New York City SPBP, which we discuss in the remainder of this report.

## The New York City Schoolwide Performance Bonus Program

The New York City SPBP was a pilot program resulting from negotiations between NYCDOE and UFT, the union that represents teach-

ers and other education employees for purposes of setting salaries and working conditions.

NYCDOE manages the largest school district in the country, serving approximately 1.1 million students in more than 1,700 schools across the five boroughs of New York City. The department is led by a chancellor, who is appointed by the mayor, and is divided into 32 community districts, each led by a superintendent overseeing the elementary and middle school students in that area. High schools are overseen by high school superintendents in each borough. In addition, the schools in each borough are served by an integrated service center, which provides administrative support in a variety of areas and 11 internal and external school support organizations, which help schools reach accountability targets, provide professional development, and help design programs and interventions.

Founded in 1960 as a local affiliate of the American Federation of Teachers, UFT represents approximately 200,000 members and is the sole bargaining agent for most of the "nonsupervisory" educators who work in the New York City public schools—including 87,000 teachers and 19,000 classroom paraprofessionals, and guidance counselors, social workers, school secretaries, attendance teachers, psychologists, adult education teachers, administrative law judges, nurses, laboratory technicians, and speech therapists.[10]

The following section draws on data gathered from written program documents distributed by NYCDOE, UFT, or both jointly, as well as from interviews conducted in 2009 with key NYCDOE and UFT policymakers involved in the design and implementation of SPBP.[11] The next sections of this chapter describe the program's history, goals and theory of action, funding, design components, timeline, and results.[12]

---

[10] UFT also represents more than 53,000 retired members, as well as teachers and employees of some private educational institutions and charter schools (UFT, 2011).

[11] See Chapter Three for more details on data sources and methods.

[12] Unless otherwise noted, all quotations in these sections come from interviews conducted for this study.

### History of the Program

SPBP's origins date back to contract negotiations between UFT and NYCDOE in 2005. At that time, the school system and the union agreed to hold future discussions on a number of topics that proved too complicated to resolve in 2005. One of these issues was an alternative form of teacher compensation—one that would either adjust or complement the single salary schedule that provided increases based on years of experience and graduate credits.

In 2007, the time was ripe to take up this issue. As one NYCDOE official commented, "The stars kind of aligned." In November 2007, an MOU between NYCDOE and UFT established SPBP:

> The New York City Department of Education and the United Federation of Teachers jointly support, and pledge together to implement on a pilot basis, a schoolwide-based bonus program pursuant to which educators will be awarded substantial bonuses for student achievement gains.

The principal negotiators of the agreement were Chris Cerf, then–deputy chancellor for Organizational Strategy, Human Capital, and External Affairs for NYCDOE, and Randi Weingarten, then-president of UFT.[13] Established as a two-year pilot for the 2007–2008 and 2008–2009 school years, the program was extended for a third school year, 2009–2010, in 2009 as a result of additional negotiations between NYCDOE and UFT.

While both parties agreed to revisit the teacher compensation issue in 2007, determining what a new compensation system might look like was less straightforward. The resulting program—SPBP—was the product of a series of school district–union compromises. The ultimate plan for SPBP was approved by a vote of UFT's delegate assembly, a representative governance body consisting of members elected from each school. UFT officials reported that the proposed compensa-

[13] Weingarten resigned as UFT president in August 2009 to assume full-time duties as president of the national American Federation of Teachers. In 2009, Cerf took a position with Mayor Michael Bloomberg's reelection campaign and, in December 2010, was appointed Commissioner of Education for the New Jersey Department of Education.

tion plan was approved overwhelmingly. Said a UFT official, "We [the UFT] told [teachers] this is a school program. Let the school decide to go in or not. If you [delegates] vote down the agreement, then you won't even give [schools] the opportunity to vote yes or no [on going into the program]. And the delegates saw it that way."

Finally, NYCDOE leaders viewed the policy as complementary to a series of concurrent reforms being undertaken districtwide under the Children First initiative. Mayor Bloomberg, with Schools Chancellor Joel Klein, launched this reform initiative in 2002 "to create a system of outstanding schools where effective teaching and learning is a reality for every teacher and child" (NYCDOE, undated). Children First began with the reorganization of NYCDOE's management structure, focusing on centralizing a decentralized system. The department also adopted a host of reform policies that included

- a systemwide approach for instruction in reading, writing, and mathematics that was reinforced with annual testing in grades 3 through 8
- a new parent support system that provided a parent coordinator for each school
- new student promotion policies
- a leadership academy to train new school leaders.

Over the next few years, the district also piloted and expanded programs to increase school autonomy in exchange for greater accountability for achievement results. In 2007, schools not meeting accountability targets faced consequences, and under the new SPBP, a subset of participating schools meeting and exceeding targets could receive financial rewards. At the public announcement of SPBP, Chancellor Joel Klein noted that the program "supports and reinforces our Children First reforms, which drive decisionmaking and resources to schools, where they can have the greatest impact on student learning" (NYCDOE, 2007).

### Goals and Theory of Action

According to both the school system and the union, the primary goal of SPBP was to improve student achievement. Both NYCDOE and UFT agreed that student learning was not what it needed to be. Both were clear that raising measurable levels of student achievement was their common, and principal, priority. As NYCDOE's Chris Cerf remarked,

> One of the barriers to successful school districts is [that] there is not a common definition or metric of success. It's very, very, very important to be unambiguously clear about what constitutes success in the system. And success for us [the NYCDOE] is measured by how much children are learning. . . . Getting children to learn more in ways that are measurable and objective is the definition of success.

Despite agreeing on the ultimate purpose of the policy, NYCDOE and UFT differed about which specific compensation-related strategies might offer the best prospects for increasing student achievement. NYCDOE favored an incentive-pay system that would reward individual teachers for their students' test score gains as the way to motivate teachers to change practices to improve student achievement. Performance pay, the department believed, was a way of highlighting teacher effectiveness. Many of the comments from NYCDOE and funders at the outset of the program emphasized the understanding that bonuses would serve as rewards and incentives that would motivate behavior (NYCDOE, 2007):

> This program will allow us to reward our most successful schools and educators. It will also help us motivate educators to work together to come up with innovative solutions to help our students and our schools excel. (Chancellor Klein)

> We know from experience across other industries and sectors that linking performance and pay is a powerful incentive. I applaud the Mayor, the Chancellor, and the UFT President Randi Weingarten for taking the bold step of applying this model to teaching as teachers are one of the most important levers in New York's overall effort to improve student achievement. (Mr. Eli Broad, funder)

UFT, while open to changing teacher compensation practices,[14] opposed the kind of individual merit pay that "put teachers in competition within a school." Instead, the union believed the key to lifting student performance lay in a system that tied financial bonuses to cooperation among teachers and between teachers and the principal. Collaborative school environments, the union reasoned, would take teachers out of their isolated classrooms and foster professional communities in which educators would work together toward common goals and outcomes. Said then–UFT President Randi Weingarten:

> We have not yet seen a reliable, careful way that isolates the effectiveness of individual teachers based upon the test scores of their students. But what we do know is there are a lot of directional things that are important. And we do know that if you end up having people really work together at a school level that does improve achievement.

Similarly, in her public statement at the announcement of the program, Weingarten commented:

> This schoolwide bonus program recognizes and builds upon the UFT's core philosophy that students learn, achieve and benefit most when all educators in a school collaborate to provide the best possible education. It properly refocuses the misguided debate over individual merit pay. Respecting and understanding the importance of teamwork and collaboration is precisely why the UFT enthusiastically supports this schoolwide initiative. (NYCDOE, 2007)

In developing SPBP, NYCDOE was mindful of the union's political challenges. Said Chris Cerf, "My goal was to find a way to . . . push for some kind of [pay] differential while having a . . . recognition of

[14] The teacher salary schedule in New York City, for example, awards additional pay to teachers who earn National Board Certification. In addition, under Rudy Crew's tenure as chancellor of New York City public schools in the late 1990s, UFT worked with NYCDOE to create the Chancellor's District, a geographically noncontiguous "district" encompassing the school system's lowest performing schools. Teachers in Chancellor's District schools were paid more than teachers in other city schools in exchange for greater accountability and a longer school day and school year.

some of the political challenges that the union would have in moving in that direction." NYCDOE leaders also recognized the value of using schools as the unit of analysis to promote collaboration. Cerf noted, "My personal philosophical inclination is that [having it be] school-based, or a substantial portion of it be school-based, is the right way to go. . . .The research is pretty clear that some kind of collaborative environment is conducive to good student outcomes."

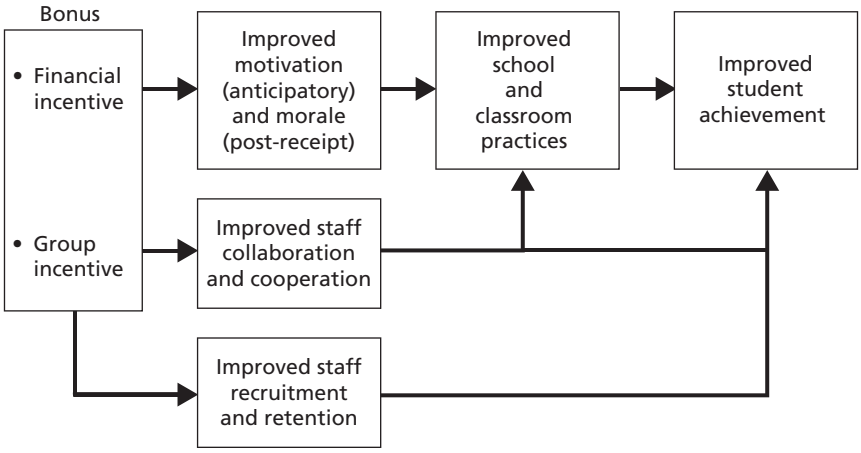
In addition to improving student achievement, the individuals who helped design and fund the program articulated a number of ancillary and intermediate goals. First, several stakeholders described SPBP as a way of both retaining high-quality teachers in hard-to-staff schools and encouraging other teachers to seek openings in these schools—which would ultimately improve the performance of students in these schools. For example, a UFT official explained, "One goal of the program would be as a recruitment and retention tool for teachers in hard-to-staff schools." A NYCDOE official concurred that, "If successful, the program would help to attract and retain effective teachers in high-needs schools." NYCDOE (2007), the official press release announcing the program, emphasized this goal: "Because only high-needs schools are eligible to participate in the program, it creates incentive for great teachers to teach in City schools that serve high-needs, low-achieving students."

Second, a representative of one of the key funders in the program's first year of implementation viewed SPBP as opening the door to broader teacher compensation reform: "A goal [of SPBP] is to push the lever a little bit in a different way as a start in teacher compensation reform."

Ultimately, NYCDOE's and UFT's differing beliefs and hypotheses about what strategies would contribute to increasing student achievement were melded into a set of compromises that became SPBP. As UFT's Weingarten put it, "We walked in their shoes, and they walked in our shoes."

As Figure 2.1 illustrates, district and UFT leaders articulated several hypotheses about how SPBP would improve student outcomes:

**Figure 2.1**
**SPBP Theory of Action**




RAND *MG1114-2.1*

- Some believed that the possibility of receiving a share of the financial bonus would motivate individual staff members to improve their efforts and practices to maximize chances of achieving that bonus, which would improve practices and, ultimately, student achievement (the anticipatory effect of the incentive). Within this group, some further believed that the possibility of schools differentiating bonus distributions based on individual performance might enhance the motivational effects of the bonus and lead to even greater positive outcomes for students.
- Others believed that winning a financial bonus would improve staff morale, which would also improve practices and student achievement (the postreceipt effect).
- Still others emphasized that providing a bonus tied to schoolwide performance would be an incentive for all staff to work together to maximize their chances of achieving that bonus, and that the improved collaboration and cooperation would improve student achievement either directly or indirectly as a result of improved school and classroom practices.

- Finally, a few leaders hypothesized that participating in the bonus program might motivate staff to seek and sustain employment at these schools, which could in turn provide greater continuity and support for school improvement efforts and ultimately improve student achievement.

### Program Funding

As indicated in the MOU, NYCDOE and UFT agreed that the school system would try to raise outside private funds to pay for the program's first year and would use public funds to sustain SPBP in subsequent years. According to NYCDOE's Chris Cerf,

> The mechanical problem we were solving for [initially] is that we [launched the program] in the middle of a fiscal year when budgets were set and there was no money. And so we agreed that we would ask funders to pay for the first year, but the public dollars would pay for ensuing years.

For SPBP year 1, the school system secured $12.76 million from private sources, including more than $5 million from the Eli and Edythe Broad Foundation, $5 million from the Robertson Foundation, and the remaining funds from other foundations and the Partnership for New York City.[15] Year 2 drew on $33 million in public funding. Year 3 drew on $4.2 million in public funding.

According to NYCDOE, UFT, and the funding partners, the funders played no role in shaping the compensation program. However, they remained invested and interested in its progress and outcomes.

### Program Design Components

New York City SPBP's design elements typified those of a school-based performance-pay program (i.e., school as unit of accountability) in which all participating schools meeting or exceeding perfor-

[15] The Partnership for New York City is a nonprofit member organization comprising 200 CEOs from New York City's corporate, investment, and entrepreneurial firms. Among the partnership's top priorities is supporting continued reform of the city's public education system (Partnership for New York City, 2011).

mance targets received a bonus (i.e., a fixed-performance approach). To receive such awards, participating schools needed to improve their performance above a certain threshold, determined by a grading system based in large part on student growth on standardized tests, as well as on attendance and school environment (i.e., a limited linear model relying on educational inputs and outputs). This subsection describes SPBP's design elements in more detail, including school eligibility and participation requirements, methods for calculating and distributing bonuses, and support and communication provided to participating schools. In all, most of these elements stayed the same over the course of the three years.

**School Eligibility for Participation.** By mutual agreement between NYCDOE and UFT, only high-needs schools were eligible to participate in SPBP. High need was defined by the factors used to create peer groups for the citywide Progress Reports (see description of these reports below). For high schools, the criteria were the average proficiency ratings for actively enrolled students on the 8th grade New York State ELA and mathematics exams. For middle schools, the criteria were the average proficiency ratings on the 4th grade ELA and mathematics tests. For elementary schools, the criteria were the poverty rate, other student demographic characteristics, and the percentages of English language learners (ELL) and special education students in the school. "There wasn't a lot of controversy about what the high-needs schools were; of course, [they were] the schools with underperforming students," remarked a UFT official.

These criteria yielded a list of 427 schools, of which about one-half were randomly selected to be invited to participate in year 1.

**School Participation Requirements.** To maximize the prospect of teacher buy-in, leaders made program participation voluntary.[16] In SPBP-eligible schools, 55 percent of the total UFT-represented employees assigned to the school—not just 55 percent of those there on the day of the election—were required to vote in the affirmative for the

[16] While participation was voluntary for the schools, it was not for individuals in the schools that agreed to participate.

school to participate.[17] The principal also had to agree. School "yes" and "no" tallies were submitted to NYCDOE and UFT. Each SPBP-eligible school was required to vote annually on whether or not to continue to participate in the program.

Said immediate past UFT president Randi Weingarten,

> The details are what created the credibility [for the program] in the schools because we [the UFT] were able credibly to say to our members that this is really up to you. You decide each year whether you want in or out.

**Criteria for Earning a Bonus: Performance Measures.** Unlike some other pay-for-performance programs that operate as rank-ordered tournaments, SPBP was based on fixed performance measures. Schools that met the target for a bonus received one. Bonuses were based on meeting the performance targets on the Progress Reports that NYCDOE issued all schools districtwide. "It's very important to have all [of] our performance metrics aligned around a common thing, just as a matter of changing the culture," said then-NYCDOE's Chris Cerf. "The Progress Reports . . . evaluate the schools, and they're a basis upon which we close schools or reward schools, so it made sense to bring [the compensation program] into alignment."

Each Progress Report—which all schools districtwide receive—compares a school's student-learning results both against the results for schools serving the same grades throughout the school system and against those for a peer group of up to 40 schools with similar student populations. Performance is based on multiple indicators, although score calculations weigh output measures more heavily than input measures. In 2008–2009, a school's overall score on the Progress Report consisted of multiple qualitative and quantitative factors, as summarized below:

---

[17] See Chapter Four for details on these annual votes. In the vast majority of schools, vote counts far surpassed the 55-percent threshold. Of the193 schools for which we have data for year 2 and the 182 schools for year 3, more than 90 percent of staff voted to participate in almost two-thirds of the schools.

- *Student performance* constitutes 25 points of the score. For elementary, middle, and K–8 schools, performance is measured by students' annual scores on the New York State tests in ELA and mathematics (median proficiency and percentage performing at or above proficiency). Student performance in high schools is measured by graduation rates (4- and 6-year graduation rates and weighted diploma rates).[18]
- *Student progress* composes 60 points of a school's score. Student progress for elementary, middle, and K–8 schools is measured by average school improvement on New York State tests from the previous year (average change and the percentage making progress[19]). For high schools, the measures are credit accumulation and completion and weighted pass rates for the Regents Examinations.
- *School environment* makes up 15 points of the Progress Report score. School environment consists of factors such as student attendance (up to 5 points) and results of NYCDOE-issued student (middle and high school), parent, and teacher surveys (up to 10 points) which measure perceptions about academic expectations, communication, engagement, and safety and respect at the school.
- *Additional credit for exemplary progress* with high-needs populations can add up to 15 points to the Progress Report score. Elementary, middle, and K–8 schools earn this credit for groups of high-needs students (ELL, special-education students, and students in the lowest one-third citywide) who demonstrate exemplary progress on test scores (defined as 1.5 years of gains on state tests). High schools earn extra points for high-needs students'

[18] NYCDOE, 2009, p. 7, expands on median proficiency:

> This measure represents the median Proficiency Rating for all students in the school. The "median" is the midpoint of all students: half of all students had a higher score; half had a lower score. As is described in the definition of Proficiency Ratings above, this Median Proficiency Rating is measured on a scale of 1.00 to 4.50, based on the scale score.

[19] Both average change and percentage making progress are based on proficiency ratings, not scale scores.

credit accumulation (i.e., earning 11 or more credits in their first, second, or third years of high school), Regents Exam scores (earning a 75 or higher for the first time on ELA or mathematics Regents), and graduation with a Regents Diploma. (NYCDOE, 2009a, c, and the Progress Reports website itself).

Final calculations of Progress Report scores are based on the school's overall points relative to all schools citywide and to schools with similar student populations. Each school receives an overall score that is a weighted average of school environment (15 percent), student performance (25 percent), and student progress (60 percent) plus any additional credit earned. Letter grades are assigned to schools based on these overall scores and scores within the three categories.

In the first two years of the program, NYCDOE determined cutoff scores corresponding to each letter grade.[20] In year 3, in an effort to increase the rigor of proficiency standards, the state department of education raised the cutoff scores for determining proficiency on state tests. This change decreased student passing rates throughout New York City and the state. To address this, in March 2010, NYCDOE announced it would change the way it determines Progress Report grades, relying on growth percentiles and a set distribution of grades (25 percent of schools would receive As, 35 percent Bs, 35 percent Cs, 4 percent Ds, and 1 percent Fs).[21] As a result of the recalibrated state test, school grades generally declined across the city.

In year 1, NYCDOE set annual performance targets based on Progress Report scores for all schools in the district.[22] Although the

[20] In the first year of Progress Reports, NYCDOE set cut scores based on the distribution of scores in that year. In subsequent years, cut scores were to remain the same, regardless of the distribution. This changed in 2010, as described above.

[21] NYCDOE, 2010c, explains this as a model that "measures the change in student test scores from last year to this year by comparing students who started at similar levels of proficiency."

[22] From its weighted total score on the Progress Report, each school was assigned to one of five percentile ranking categories based on the range of all weighted total scores citywide, by type. These ranks were used to determine each school's target for the next year. A school in the 85th percentile rank or above was given a target of gaining at least 7.5 points; from the

Progress Reports scores compared schools to others in their peer group and to all schools in the district, the performance targets were not explicitly normative. Every school had a specific target for its Progress Report score and grade that was independent of the targets for other schools. Moreover, if all schools met their targets, all schools could earn bonuses. Eligible schools that met their targets by 100 percent or exceeded them received the full bonus, equal to $3,000 multiplied by the number of full-time UFT-represented employees in the school. As a result of a modification to the NYCDOE-UFT MOU in the middle of year 1, eligible schools that met at least 75 percent of their targets received one-half of a bonus pool equal to $1,500 per UFT-represented staff member. For example, if a school had a Progress Report score of 60 in 2007–2008, a target of 68 in 2008–2009, and earned a 66, that 6-point increase represented 75 percent of the 8 points needed to reach the 2008–2009 target. Therefore, this school would have earned one-half of the full bonus amount.

At the end of year 1 (June 2008), NYCDOE and UFT announced an additional means by which a participating school could earn a partial bonus: maintaining an "A" Progress Report grade for two consecutive years. District leaders believed this adjustment was necessary for schools that had already achieved a high Progress Report grade and had less room for growth than other schools did but still deserved an opportunity to earn a bonus. Additionally, a small number of schools participating in SPBP had Progress Report scores greater than 92.5, which meant these schools could not meet their performance standards unless they earned points through additional credit.

In year 2, NYCDOE eliminated targets for all schools except for those participating in SPBP. For these, the department adopted new targets specifically for determining bonus recipients (using the same formula as in the previous year). These targets were "intended to incentivize schools to move up at least one grade on the Progress Report" (NYCDOE, 2009a). Schools with lower Progress Report scores in the

84th to 45th percentile, the target was at least 12.5 points; from the 44th to 15th percentile, at least 15 points; from the 14th to 5th percentile, 17.5 points; and below the 5th, percentile 20 points.

previous year were required to make greater improvements to reach their targets. For example, in 2008–2009 (year 2), elementary, middle, and K–8 schools receiving an F the previous year needed to rise to a C (i.e., a minimum score of 43) or increase by at least 20 points (whichever results in a higher score); D schools needed to rise to a C or least 17.5 points (whichever was greater); C schools needed to rise to a B (i.e., score of 54) or go up at least 15 points (whichever was greater); B schools were required to go up at least 12.5 points and A schools at least 7.5 points. For the elementary, middle, and K–8 schools, NYCDOE set 75.5 as the maximum target—or the minimum score to receive an A (68 points) plus 7.5 points.[23]

During year 3, NYCDOE considered altering the criteria for SPBP bonuses, in anticipation of changing proficiency standards for the state test and potential declines in school-level performance and, therefore, eligibility for bonuses. The department could not, however, reach an agreement with UFT on how to do so. The resulting indecision about how targets would be set or how bonuses would be determined continued through the academic year, leaving schools without information on bonus criteria. Ultimately, NYCDOE decided to use the same methodology as in year 2. As discussed later, this meant that far fewer schools received bonuses for the 2009–2010 school year.

**Distributing Bonuses.** Bonus money, given to a school in a lump sum, was distributed within a school by a school-based CC. The four-member committee at each school consisted of the principal, another individual chosen by the principal, and two members selected by UFT-represented staff at the school.[24] The CC had full authority to allocate the school's bonus. It could divide the money equally among all eligible school recipients or could differentiate dollar amounts, giving larger sums to some individuals than to others. The one restriction was that the committee could not award money on the basis of seniority. While

[23] A school receiving two Progress Reports (e.g., a K–12 school receiving an elementary and high school report) qualified for a bonus if it (1) met at least 75 or 100 percent of one of its targets and (2) did not receive a lower score in that year than in the previous year on the other report.

[24] In the 2008–2009 school year, two schools reported to NYCDOE that their CC had only three members. We also heard from a school whose CC had five members.

the formal MOU did not cover school staff members not represented by UFT, CCs were free to decide to allocate bonus money to these personnel. Individual recipients within a school could also opt to donate their shares or some portion of them to the school, but again, this was not a formal part of the program.

According to UFT and NYCDOE leaders, CCs were incorporated into the design of SPBP to resolve two potential dilemmas the program might otherwise have confronted. First, both NYCDOE and UFT wanted to make sure that decisions about awarding bonuses were made at the school, not centrally by the school system or the union. As one UFT official said, "The schools were the unit that was opting into the program, so it [made] sense to have the decision [about how to distribute the bonus] at the school level." Second, NYCDOE wanted bonuses to be differentiated by job title; the union did not. "Leaving it up to each school as far as how the money was allocated . . . avoided that problem," said a UFT official.

As former UFT president Randi Weingarten commented, "[Chris] Cerf and I kept wrestling back and forth with, 'How do you really assure that these decisions [about awarding bonuses] will come from the schools?'" The "conceptual breakthrough," according to NYCDOE official Chris Cerf, was the CC. CCs would ensure that decisions about how to distribute bonuses were school-based. Moreover, explained Cerf,

> If a Compensation Committee [chose] to award people differentially based on whatever considerations they thought [made sense], whether they had different jobs in the building or they had different levels of performance, they could do that. If, on the other hand, they felt that the value of share and share alike [held], if there [was] collective success for the building, they could [award everyone the same amount].

According to program guidelines, the CC was required to make its decisions by consensus. No votes were supposed to be taken, so neither the administration nor the UFT members could control the results (equal representation of administrator and UFT appointed members was also intended to serve this purpose). If the CC could not reach consensus on how to allocate the bonuses, the school was to for-

feit any money it would have been awarded. Thus, the committee had a compelling incentive to come to agreement about how to distribute bonus money the school had earned.

The CC determined the school's bonus distribution plan in the spring and specified which staff members would receive bonuses and how much each would receive when the NCYDOE announced bonus results in the subsequent school year. These decisions were made before test scores were released and before NYCDOE determined which participating schools had earned bonuses. According to one NYCDOE official, the decision to determine bonuses before test scores were released was purely practical. When SPBP was first established, state tests took place too late for them to be considered in committee decisions because results were not released until the summer. Other district leaders reported struggling about when to decide on distribution plans. On the one hand, they recognized that figuring out distribution plans prior to the release of test results would limit the option of using approaches that differentiated based on performance. On the other hand, they feared that determining distribution plans after the release of test results would weaken incentives for teachers and schools to improve performance, given that they would feel that they were working toward an unknown or moving target.

Ultimately, bonuses were awarded in the fall of the next school year (in year 3, the announcement came much later) and followed individual staff members to the school in which they currently worked. So, if a teacher moved to a new school that fall, he or she would have received payment for the bonus earned by the school in which he or she taught the previous year. The chancellor of the school system and the president of UFT constituted an oversight committee to review any appeals of bonuses.

**The Bonus Share Amount.** NYCDOE and UFT arrived at a bonus share of $3,000 per UFT-represented employee in an attempt to strike a balance between the financial circumstances of NYCDOE and ensuring that compensation bonuses were large enough to attract teachers' interest and participation. NYCDOE official Cerf explained:

> [We] looked at the literature . . . and that number [$3,000] . . . is [about] five percent of [the salary of] a fourth-year or fifth-year teacher. And that's all we could afford. In an ideal world, it would have been more. I think there's a little bit of literature about this that says to really change behavior—and that's the only reason you would do something like this—that number needs to be bigger. But [we] thought intuitively that that number was big enough to have some meaning.

UFT's Weingarten echoed Cerf's explanation:

> That was actually the minimal amount we thought was necessary to spark people's interest. Of course, it should be higher. But we thought that given . . . the other needs of the system, and particularly walking into a fiscal crisis, you had to be very careful not to make this experiment more important than so many other things that have to be done.

**Communicating to Schools about SPBP.** SPBP MOU specified that NYCDOE and UFT issue joint communications to schools about the program. This requirement was designed to emphasize that the program was a partnership between NYCDOE and UFT and to ensure that all school personnel received the same message (see next section on "no lobbying").

The program was initially announced with a joint letter from the Chancellor and the UFT President. Joint NYCDOE-UFT orientation sessions were conducted for eligible schools. Written information about the program, such as general descriptive bulletins, was also issued jointly. One UFT official noted, "We [the NYCDOE and UFT] [do] joint orientations and joint releases to the schools. We really [try] to work together to have this plan succeed."

As discussed in Chapter Four, lessons learned from SPBP's year 1 orientations were used to improve year 2 orientations. For example, in year 2, NYCDOE and UFT required participating schools to attend the orientation sessions. Also, NYCDOE developed and made available to all schools in the district the Progress Report "Modeler," a software tool to help schools examine their Progress Report scores and

target, isolate the various components (e.g., student performance, student progress), and identify what outcomes they needed to achieve to reach the targets or improve their scores. At least one NYCDOE leader interviewed believed this tool could assist participating schools interested in strategizing to improve their chances of earning the bonus.

**No-Lobbying Agreement.** A significant feature of the NYCDOE-UFT SPBP agreement was that both sides encouraged eligible schools to participate in the program but that neither side would attempt to lobby schools. In the words of the MOU, "The NYCDOE and UFT jointly pledge to work in good faith to encourage [eligible schools] to participate . . . in the first year and throughout the life of the program."

According to both NYCDOE and UFT officials, each side kept its word. Said a NYCDOE official, "The UFT were honest brokers on this. We had a series of meetings [at schools] with a NYCDOE representative and a UFT representative in which we both describe the program and I think passively advocate for its acceptance." A UFT official concurred: "We did not pressure any school to go into [the program]. We just gave them the information." Noted another NYCDOE official,

> I think it [is] the clear message from UFT and DOE that this was a worthwhile program to participate in. So we didn't go out and lobby schools . . . but we put out positive messages about the program. . . . The message from the UFT . . . to the schools was that this is a good way to test the power of collaboration. And from us [NYCDOE], it was that it was a good way to reward groups of educators doing an effective job with kids. And I think the subtext from both of us was that there's not a real downside to participating.

This supportive but relatively neutral stance extended to the issue of advising CCs on how to distribute bonuses. A UFT official explained, "I was very clear at all of the orientations. I said, 'This is your program, we're not going to tell you how to [distribute the bonuses]. You do what you think is right.'" As a NYCDOE official put it, "They [the schools] were totally on their own. . . .A formal part of the deal is that both the UFT and the NYCDOE expressly agree not to try to put a thumb on the scale."

**Alignment with the Administrators' Bonus Program.** To maximize SPBP's chance to be successful, leaders felt strongly that principals had to support it. As a result, program guidelines dictated that, in addition to the requisite agreement of 55 percent of the staff to participate, the school's principal also had to concur. To enhance principals' commitments to SPBP, NYCDOE designed a targeted incentive for them. The department hoped that linking SPBP to a bonus for principals and assistant principals would give administrators even more incentive to work with teachers to raise test scores.

Even prior to SPBP, New York City principals had a financial incentive program. As a result of an agreement between NYCDOE and the Council of School Supervisors and Administrators (CSA),[25] principals whose Progress Report scores are in the top 20 percent citywide among all public schools are eligible for a $25,000 bonus (assistant principals are eligible for one-half that amount). Unlike SPBP, this program is competitive; only a small number of administrators (those whose scores are in the top 20 percent) earn the bonus. According to a NYCDOE official, since few principals are likely to earn the $25,000 bonus,

> [we] realized that principals who were in schools where people were eligible for [SPBP] bonuses had no stake in the program. We needed to make sure that the principal bonus plan and the teacher bonus plan dovetail. We couldn't have a situation where teachers [are] earning a bonus but principals were not. So we created a kind of catch-all tier in the principal bonus program so that wouldn't happen. We came up with a mechanism to make sure [principals] got a bonus if the teachers got a bonus.

That mechanism was to award $7,000 to principals in schools that received the full SPBP bonus and $3,500 to principals in schools that received one-half the SPBP bonus. Assistant principals earned one-half the amount awarded to principals in both cases. Administrators who qualified for both the CSA and the SPBP bonus in one year only

[25] CSA bargains collectively for New York City's principals, assistant principals, supervisors, and education administrators.

received the CSA bonus. According to UFT, "That [arrangement] gives [principals] an incentive to want to see the [SPBP] succeed."

### Implementation Timeline

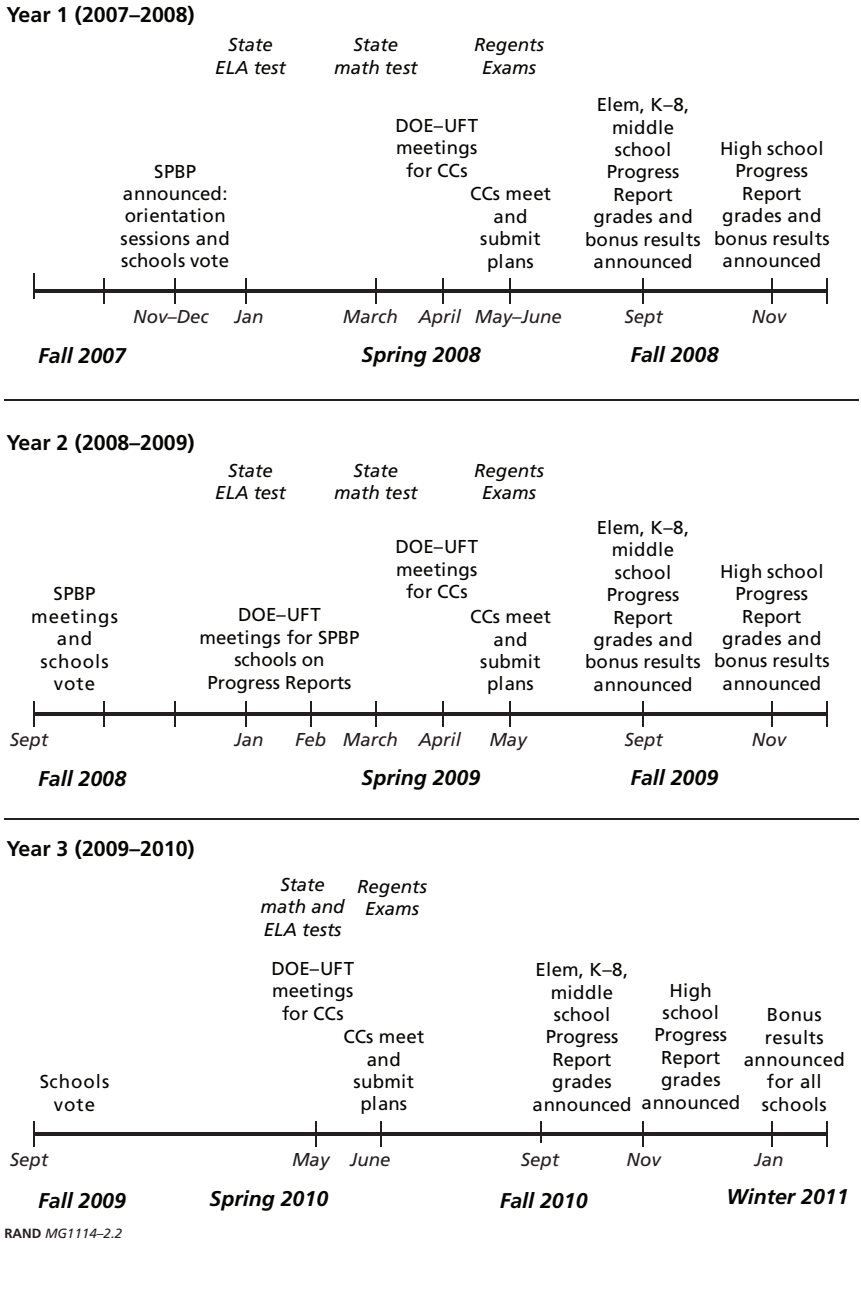
Figure 2.2 illustrates the timeline for SPBP implementation during its three-year existence. In year 1, the program was announced in October, and schools were asked to vote to participate in November and December 2007. Following a series of training sessions in April and May 2008, CCs were then asked to meet and submit a distribution plan to NYCDOE in spring 2008. In fall 2008 (the subsequent school year), year 1 Progress Report grades and bonuses for elementary, K–8, and middle schools were announced. In November, the high school Progress Report grades and bonuses were announced.

In year 2, schools were notified about the program at the start of the school year and were required to vote in September 2008. In January and February 2009, NYCDOE and UFT also held meetings for CC members to disseminate information about the Progress Report criteria. Following a series of training sessions in April and May, CCs were asked to meet and submit a distribution plan by the end of May 2009. In fall 2009 (the subsequent school year), year 2 bonuses were announced for elementary, K–8, and middle schools. In November, the high school bonuses were announced. In year 3, schools were notified and required to vote at the start of the year. Program leaders believed that, because no new schools were participating, they did not need to hold additional meetings at the start of the school year. Leaders believed most schools were fairly well oriented to the program and the Progress Report system. In early May, leaders once again invited CC members to attend informational meetings and submit their distribution plans by the end of May (some follow-up occurred in early June). The year 3 leaders delayed the announcement of bonuses for elementary, K–8, middle, and high schools until January 2011.

In year 1 (2007–2008), 427 high-needs schools were identified for the program.[26] Originally, a phased rollout was planned, but this

[26] Fryer (2011) reports that 430 schools were initially identified for the study. However, there was considerable ambiguity about the schools in the documents we received from

**Figure 2.2**
**Timeline of SPBP Implementation**




RAND *MG1114–2.2*

never occurred. Instead, a lottery was used to invite about one-half of the identified schools to participate in SPBP. Roland Fryer, Harvard University Professor of Economics and former Chief Equality Officer with NYCDOE (2007–2008), was responsible for the selection and random assignment process. Lottery selection process files provided by NYCDOE identified the schools participating in the lottery assignment process. According to those files, 25 of the high schools identified as high-needs schools were removed from the list prior to the lottery.[27] Initially, the department used a random-number generator to select a sample of the remaining 402 for invitations to participate in the program. When several of the schools declined the invitation to participate in the program, a second small sample of schools was selected to receive invitations to participate in SPBP from the sample of schools originally assigned to control.[28] The two-stage lottery thus netted a total of 234 invitations to randomly selected high-needs schools, while 168 others were not selected.

In these schools, 55 percent of the UFT-represented staff had to agree to participate in SPBP. Thirty-two schools that were invited to participate in SPBP voted not to participate during year 1 (2007–2008), and three schools withdrew, two after initially voting in favor

---

NYCDOE. The department confirmed that 402 schools were randomly selected in the lottery, and reports suggest that 25 were removed prior to randomization, for a total of 427 schools initially identified. Our sample includes 234 schools that were randomly selected to receive an invitation to SPBP and 168 schools that were randomly selected for the control group. The sample included one school that was randomized to receive an invitation to SPBP but that was not included in the sample used in Fryer (2011). We received confirmation from NYCDOE that the school took part in the lottery and was assigned to the SPBP group. Similarly, our sample includes five schools that NYCDOE confirmed as taking part in the lottery and being assigned to control but that were not part of the sample used in Fryer (2011).

[27] We have very limited information on school removal decisions and cannot explain the exact reasons for the removal of these 25 schools from the initial list.

[28] Data from the district and the Fryer (2011) report both indicate that 21 schools were selected for the second sample, which implies that 213 schools received the initial invitations to participate in SPBP. However, we have no data specifically identifying the schools in the initial sample, and Fryer (2011) suggests that only 212 schools were in the original sample selected for SPBP. As noted above, our sample includes one additional school in the sample chosen by lottery to receive an invitation to participate in SPBP, which is consistent with the discrepancy between the sample sizes for the original samples.

of participation and one that could not reach a consensus vote and was therefore excluded. In the end, 199 of the schools selected by the lottery participated in SPBP in year 1.

According to the lottery selection files, eight schools (mostly special education schools) not included in the lottery were also invited to participate in SPBP. Six of these schools participated, and two did not. The final total of schools participating in the program in year 1 was thus 205.

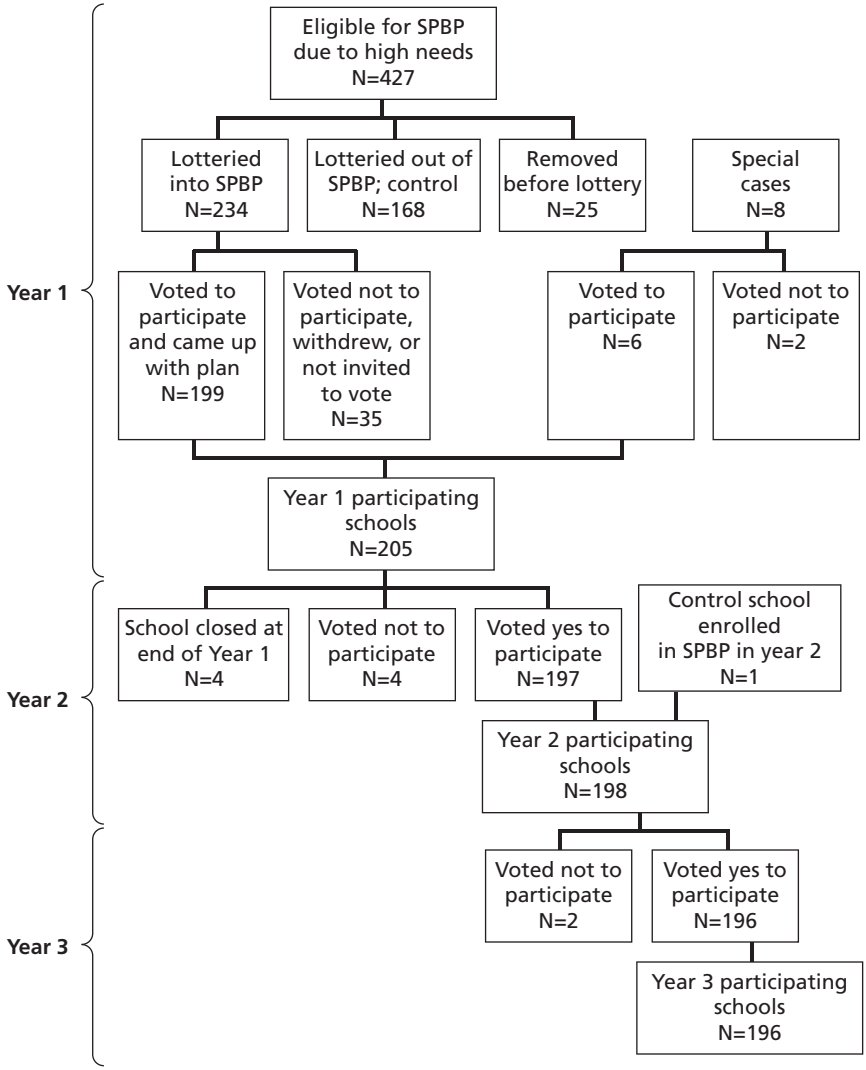
In year 2 (2008–2009), 198 schools participated. Four schools dropped out of the program because they failed to vote or voted not to participate; four schools were closed (all from the sample of schools selected by lottery to receive an invitation to participiate in SPBP); and one control school was enrolled in the program per negotiations with that school, reflecting a net loss of seven schools from year 1 to year 2. In year 3 (2009–2010), an additional two schools (again both from the initial lottery-chosen set) voted not to participate in the program, with 196 schools continuing to participate. Figure 2.3 shows how schools got into the program and how the number of participating schools changed over time. Table 2.1 shows participating schools by school type and year. [29]

### Bonus Results

In year 1, 62 percent of participating schools won bonuses (full or half). High schools and schools with both middle and high grades were the most likely to win full or partial bonuses (89 percent did), while only 40 percent of middle schools won a bonus in year 1. Of the 31 schools that won partial bonuses in year 1, 20 (65 percent) attained this by meeting 75 percent of their Progress Report targets and 11 (35 percent) by earning consecutive A grades.

[29] Table 2.1 lists transfer schools as a separate category. Transfer schools are small, academically rigorous, full-time high schools designed to reengage students who are behind in high school or have dropped out. See NYCDOE, 2011a, for details.

**Figure 2.3**
**SPBP Assignment and Participation by Year**




RAND *MG1114-2.3*

**Table 2.1**
**SPBP School Participation by School Type and Year**

| | Year 1 (2007–2008) | | Year 2 (2008–2009) | | Year 3 (2009–2010) | |
|---|---|---|---|---|---|---|
| **School Type** | **N** | **%** | **N** | **%** | **N** | **%** |
| Elementary School | 96 | 47 | 92 | 46 | 91 | 46 |
| Middle School (MS) | 40 | 19 | 39 | 20 | 39 | 20 |
| High School (HS) | 35 | 17 | 34 | 17 | 34 | 17 |
| K–8 | 22 | 11 | 21 | 11 | 21 | 11 |
| HS/MS | 9 | 4 | 9 | 5 | 8 | 4 |
| Transfer School | 3 | 1 | 3 | 2 | 3 | 2 |
| Total | 205 | | 198 | | 196 | |

The number of schools winning bonuses increased in year 2, particularly in schools serving lower grades.[30] In year 2, 84 percent of participating schools won bonuses (full or half). Almost all the elementary (91 percent), middle (95 percent), and K–8 (86 percent) schools won full or partial bonuses. In contrast, 59 percent of high schools won full or partial bonuses. In year 2, all schools winning partial bonuses did so by meeting 75 percent of their Progress Report targets (i.e., the A-A rule was not invoked). In year 3, far fewer schools won bonuses than in the first two years of the program, largely because of changes in state proficiency standards. Just 26 schools (about 13 percent of all participating schools) won bonuses. High schools and schools with both middle and high grades fared somewhat better than other types of schools, with 35 percent and 38 percent, respectively, winning bonuses. Only 11 percent of elementary schools and no middle schools won bonuses. Table 2.2 summarizes all these results.

[30] This reflects the dramatic increase in the overall Progress Report scores that schools throughout the district received as the result of an unexpected increase in state test scores statewide. In fact, 97 percent of New York City's elementary and middle schools earned an A or B, while only two schools earned an F; grades fell for 40 of the 1,058 schools graded. As a result of these increases, the state decided to raise proficiency standards.

**Table 2.2**
**SPBP Bonus Results by School Type and Year**

| | Year 1 (2007–2008) | | | | | | Year 2 (2008–2009) | | | | | | Year 3 (2009–2010) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Won full bonus | | Won partial bonus | | Did not win bonus | | Won full bonus | | Won partial bonus | | Did not win bonus | | Won full bonus | | Won partial bonus | | Did not win bonus | |
| School Type | No. | % | No. | % | No. | % | No. | % | No. | % | No. | % | No. | % | No. | % | No. | % |
| Elementary | 39 | 40 | 19 | 20 | 38 | 40 | 80 | 87 | 4 | 4 | 8 | 9 | 10 | 11 | 0 | 0 | 81 | 89 |
| Middle (MS) | 12 | 30 | 4 | 10 | 24 | 60 | 37 | 95 | 0 | 0 | 2 | 5 | 0 | 0 | 0 | 0 | 39 | 100 |
| High (HS) | 30 | 86 | 1 | 3 | 4 | 11 | 17 | 50 | 3 | 9 | 14 | 41 | 12 | 35 | 0 | 0 | 22 | 65 |
| K–8 | 9 | 41 | 4 | 18 | 9 | 41 | 17 | 81 | 1 | 5 | 3 | 14 | 1 | 5 | 0 | 0 | 20 | 95 |
| HS/MS | 5 | 56 | 3 | 33 | 1 | 11 | 5 | 56 | 0 | 0 | 4 | 44 | 3 | 38 | 0 | 0 | 5 | 63 |
| Transfer | 1 | 33 | 0 | 0 | 2 | 67 | 2 | 67 | 0 | 0 | 1 | 33 | 0 | 0 | 0 | 0 | 3 | 100 |
| Total | 96 | 47 | 31 | 15 | 78 | 38 | 158 | 80 | 8 | 4 | 32 | 16 | 26 | 13 | 0 | 0 | 170 | 87 |

As Table 2.3 illustrates, the amount of SPBP bonus funds awarded increased from year 1 to year 2. In total, more than $20 million were awarded to schools in year 1 and more than $30 million in year 2. Because so few schools won bonuses in year 3, a little over $4 million were awarded in this year.

## Summary

Although pay-for-performance programs are not new to the education sector, they have become increasingly popular in recent years because of new high-stakes accountability policies and new political and financial support. These programs vary along several dimensions, including the unit of accountability, the incentive structure, the criteria for awarding bonuses, performance standards and thresholds, size of bonuses, distribution methods, and payout frequencies. Past research finds mixed evidence of the motivational effects of school-based bonus programs and indicates that motivation is often mediated by perceptions of fairness and the size of the bonus. Research is also inconclusive about the effects of performance-based incentives on staff collaboration, and some of the broader accountability literature suggests some potential desirable and undesirable effects on classroom practices. There is also

**Table 2.3**
**Total SPBP Bonuses Awarded, by School Type and Year (actual $)**

| School Type | Year 1 (2007–2008) | Year 2 (2008–2009) | Year 3 (20079–2010) |
|---|---|---|---|
| Elementary | 9,640,500 | 17,353,500 | 1,741,666 |
| Middle (MS) | 2,797,500 | 6,441,000 | 0 |
| High (HS) and transfer | 6,066,000 | 3,106,500 | 1,548,103 |
| K–8 | 1,810,500 | 3,328,500 | 84,000 |
| HS/MS | 492,000 | 399,000 | 784,595 |
| Total | 20,806,500 | 30,628,500 | 4,158,364 |

limited and mixed research evidence on how these programs affect student achievement.

The New York City SPBP was a school-based performance-pay program. Each school required 55-percent approval of its UFT-represented staff members to participate in the program. A school could earn a bonus of up to $3,000 for each of its full-time UFT-represented staff members if the school met its annual performance target, as defined by the NYCDOE accountability program. SPBP required participating schools to establish a four-person CC to determine how to distribute the bonus to staff members. The committee was made up of the principal, a member designated by the principal, and two UFT-represented staff members elected by their peers.

SPBP represented a partnership between NYCDOE and UFT and a blending of their different perspectives and theories of action regarding incentive pay. The primary goal of the program was to improve student achievement. The hypothesized mechanisms for achieving that goal included offering individual bonuses to motivate and reward teacher practices and offering incentives tied to school-level performance to ensure that staff members worked together toward common goals.

Program implementation was truncated in year 1, leaving only a few months between the announcement of the program and state testing. In years 1 and 2, the schools were notified and required to vote at the start of the year. In year 1, 427 high-needs schools were identified, and about one-half of these schools were randomly selected and offered the opportunity to participate in SPBP. Ultimately, 205 schools participated in the program in year 1, 198 schools in the year 2, and 196 in year 3. The number of schools receiving bonuses increased dramatically over the first two years: 62 percent of participating schools won a bonus (full or half) in year 1 and 84 percent in year 2. This translates to schools receiving more than $20 million in year 1 and more than $30 million in year 2. As a result of the recalibrated proficiency cut scores for the state test, school grades generally declined across the city in year 3, leading to significant decreases in the number of schools receiving schoolwide bonuses. Only 13 percent of schools received bonuses, totaling $4.2 million.

CHAPTER THREE

# Research Methods

In this chapter, we describe the research questions and conceptual framework guiding our research, along with the methodology we used to collect and analyze data on the implementation and effects of SPBP. For our analyses on the tests of the program effects on teacher and student outcomes we provide an overview of the methods and data in this chapter, and provide full details in the relevant chapters and appendices.

## Research Questions

The evaluation was designed to answer the following questions:[1]

1. How was the program implemented?
   a. What was the level of awareness of, support for, and attitudes about the program?
   b. What were the staff members' attitudes about the financial bonuses?

[1] After our interactions with the district and initial reviews of the program, we reformulated the questions NYCDOE originally posed to arrive at the specific questions addressed in this report. The following were the original questions: What are issues around implementation and communication that could be improved in future years? To what extent has the program affected the interactions between and among teachers, supervisors (e.g., principals), and students in participating schools? Has it increased levels of collaboration between supervisors and teachers? To what extent has the program led to school-level changes (programmatic, instructional, operational) that might impact student performance? Do participating schools perform better on standardized achievement tests and other bonus metrics than non participating schools and, if so, by how much?

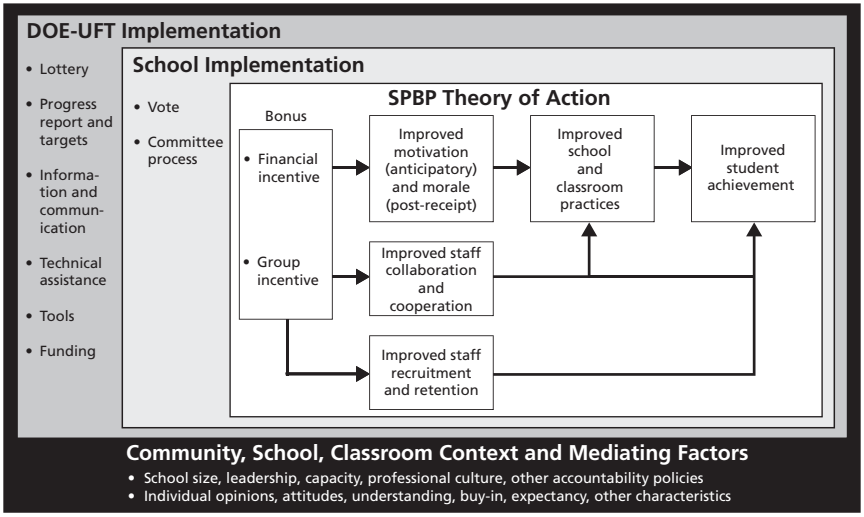
   c. What was the nature of communication about and the level of understanding of the program?
   d. How did the CC function?
   e. What were the results of the CC process? How did schools decide to distribute the bonus?
   f. How did the staff perceive SPBP's effects on school improvement and on individuals?
2. What were the intermediate outcomes of the program?
   a. How did SPBP affect teachers' self-reported attitudes and behaviors?
3. How did the program affect student performance?
   a. How did SPBP affect Progress Report scores across all years and all the component scores?
   b. How did SPBP affect student test scores across years?

## Conceptual Framework

To help answer these questions, our study design, data collection, and analysis were guided by a conceptual framework (Figure 3.1) grounded in the research on performance incentives and the NYCDOE and UFT's implicit theory of action we deduced from the interviews and review of documents.

As described in Chapter Two, district and UFT leaders articulated several hypotheses about how SPBP would improve student outcomes (inner box in Figure 3.1, labeled "SPBP Theory of Action"). Some identified the anticipatory effects of a financial incentive, believing that the possibility of receiving a financial bonus would motivate individual staff members to improve their efforts and practices in the hopes of maximizing their chances to win the bonus. Others believed that winning a financial bonus would improve staff morale, which would in turn improve practices and student achievement (the postreceipt effect). Still others hypothesized that tying bonuses to schoolwide performance would encourage all staff to work together to achieve that reward, and that this improved collaboration would improve school and classroom practices, thus directly or indirectly improving stu-

**Figure 3.1**
**Conceptual Framework**




RAND *MG1114-3.1*

dent achievement. Finally, some also argued that participation would motivate staff to seek and sustain employment at these schools, which would provide greater continuity and support for school improvement.

The framework recognizes that NYCDOE and UFT shaped the implementation process by selecting participating schools (i.e., the lottery or random assignment of schools), developing the measures through which schools could earn the bonus (i.e., Progress Report scores and targets), disseminating information about the program and how it worked, providing technical assistance (e.g., holding orientations, responding to questions and calls) and tools (e.g., Progress Report Modeler) to schools, and securing funding for the program. Schools also influenced implementation by obtaining votes and support from staff to participate and by forming committees that met to develop distribution plans.

Finally, our framework recognizes that the bonus program, like all programs, was embedded in a broader context that could influence the program's implementation and effects. As past research and theory suggest, many of these factors can mediate how individuals respond to

bonuses and affect the hypothesized causal pathways in the theory of action. For example, school size could have affected staff responses to the program and the effectiveness of the bonus incentive. Individual performance would have had a relatively less direct effect on overall school performance in larger schools than in smaller schools. As a result, it may have been easier for staff in larger schools to "free ride" (i.e., not exert any additional effort and still benefit from the bonus), compared to smaller schools in which there may be strong incentives that shape individual behavior in ways that mirror what might occur under a individual-level pay-for-performance program. As past research indicates, school leadership and capacity could also affect the level of attention paid to the bonus and efforts undertaken to attain it. Staff opinions and attitudes are also likely to mediate how individuals respond to bonus incentives. For example, staff awareness of the purpose of the program, understanding of how performance would be measured, and attitudes toward the program (e.g., perception of money as a motivating factor, perceived fairness of the method for awarding bonuses, perceived attainability of performance targets) could influence whether staff would have been motivated by the program and bonus.

## Data Collection and Analysis

The evaluation collected a variety of qualitative and quantitative data to examine the implementation and effects of SPBP. The major data collection and analysis activities we drew on for this report included surveys of CC members and teachers; site visits to SPBP schools; interviews with representatives from NYCDOE, UFT, funders, and other leaders; document reviews; meeting observations; analysis of administrative data (participation, distribution plans, bonuses); and analysis of student achievement in SPBP and control schools. We describe these activities in more detail below.

### Compensation Committee Survey

To understand perceptions about SPBP program design, implementation, and effects from the perspective of school staff closely involved with the program, we surveyed all members of the CCs in the 198

schools participating in the 2008–2009 school year and the 196 schools participating in 2009–2010.

**Survey Instrument.** In developing the survey during year 1, we drew on and adapted items from surveys previously administered by the RAND Corporation and Vanderbilt University to evaluate performance-pay programs and numerous other educational interventions (e.g., surveys administered for the evaluation of the TEEG program and the Project on Incentives in Teaching experiment in Nashville, Tennessee). We also developed many new items tailored to SPBP specifics. Colleagues within RAND and Vanderbilt, as well as representatives of NYCDOE and UFT, reviewed the draft instruments. We also pilot tested the draft survey with two principals and three CC members from two SPBP schools, and conducted cognitive interviews to obtain their feedback on the clarity of the items. The survey was revised based on the reviews and analysis of pilot testing. We developed two survey formats with equivalent questions: a paper version and an online version using SurveyMonkey. The 2010 version of the survey was similar to the 2009 version, but with a few adjustments. In some cases, new items from the teacher survey were added to the CC survey so that we could compare the responses of teachers and CC members (e.g., a new school climate measure). In other cases, we added items to collect data on domains found to be important in our first year of data collection and analysis.

**Survey Administration.** The 2009 and 2010 surveys were fielded in the final months of the school year (between April 30 and July 3 in 2009 and between May 1 and July 5 in 2010) and were administered in two phases. First, committee members attending meetings jointly sponsored by NYCDOE and UFT in late April and early May were asked to fill out paper versions of the survey during the first 30 minutes of the meeting. Completed surveys were collected by RAND staff at the meeting, and postage-paid reply envelopes were provided to some staff who were unable to complete the survey in the time given.

Later that month, we launched the online version of the survey. Because of delays in obtaining email addresses of committee members, we sent out notices in two phases. In early May, we sent principals in all participating schools email messages with the web link and asked them

to fill out the survey if they had not already done so at the meetings. Later in the month, NYCDOE staff provided us with a list of email addresses for all CC members across the participating schools, and we sent out email messages the next day with the web link and a request for participation.

All committee members had the option of requesting paper versions of the survey and, when requested, received the paper surveys and postage-paid envelopes in which to return them. Several RAND follow-up email messages were sent encouraging participation. NCYDOE and UFT leaders also sent out email messages encouraging participation.

**Response Rates.** In 2009, we received CC surveys from 606 respondents. After cleaning the data (e.g., removing duplicates, ineligibles, respondents from schools that did not participate in SPBP, and respondents who left the majority of the survey blank[2]), there were 537 respondents, for a total response rate of 68 percent. Among these respondents, we were unable to obtain school identifiers for 15.[3]

A little more than one-half of all respondents were teachers or other school staff (58 percent); the remaining respondents were administrators (127 principals and 95 other administrators; 42 percent overall). Judging by NYCDOE records and the assumption that all those identifying themselves as principals or other administrators were CSA members and all those identifying themselves as teachers and other school staff were UFT-represented staff members,[4] we obtained surveys from 70 percent of all UFT-represented members and 64 percent of all

---

[2] Because of a miscommunication, the original email inadvertently went out to some nonparticipating schools. These "ineligibles" were subsequently removed from all future follow-up emails.

[3] These survey responses were included in the analysis and results reported here but could not be factored into school-level response rates.

[4] Although SPBP was designed to include two UFT members, the principal, and a principal-designee on each committee, not all principals appointed another administrator or CSA member, as some may have expected. In fact, the percentage of teachers and other staff (i.e., UFT members) among the committee members principals appointed was approximately 30 percent. This explains why, in the aggregate, the split was not 50-50 and why more UFT members than CSA members served on the committees.

administrators serving on the committees in all participating schools. Among all respondents, 169 (31 percent) completed the paper surveys at the New York City meetings, and 368 (69 percent) completed surveys online or mailed in paper surveys later. Among all survey results received, 185 (34 percent) were hard copies, and the remaining 352 (66 percent) were generated online.[5] As illustrated in Table 3.1, administrators were much more likely to fill out the survey online.

In 2010, we received surveys from 610 respondents. After removing duplicates, ineligibles, and respondents who left the majority of the survey blank, 561 respondents remained. The total response rate was 72 percent. Among them, we were unable to obtain school identifiers for 2 respondents. About 60 percent of the respondents (332) were

**Table 3.1**
**Profile of Compensation Committee Member Survey Respondents, by Role and Mode**

| | Respondents | | Response Rate (%) | Online Respondents | |
|---|---|---|---|---|---|
| | Expected | Actual | | N | % |
| 2008–2009 | | | | | |
| Administrators | 346 | 222 | 64 | 182 | 82 |
| UFT Staff | 444 | 312 | 70 | 168 | 54 |
| Total | 790[a] | 537[b] | 68 | 352[b] | 66 |
| 2009–2010 | | | | | |
| Administrators | 345 | 226 | 66 | 177 | 78 |
| UFT Staff | 437 | 332 | 76 | 194 | 58 |
| Total | 783[c] | 561[b] | 72 | 373[b] | 66 |

[a] Based on official data reported by schools to NYCDOE via Galaxy data systems. In 2008–2009, two schools reported only 3 committee members.

[b] The totals include a few respondents who did not report their roles.

[c] In 2009–2010, one school reported only 3 committee members.

[5] Although we found some significant differences in results by survey mode (online rather than on paper) and completion scenario (during rather than after a meeting), most of these differences disappeared after controlling for respondents' roles (i.e., administrators or UFT members).

teachers or other school staff members, and the remaining 40 percent were principals (121) and school administrators (104). The respondents represented 76 percent of all UFT-represented members and 66 percent of all administrators serving on the committees in the 196 participating schools. Sixty-six percent (373) of the respondents completed the survey online. Similar to 2008–2009, administrators were more likely to complete the survey online than were UFT members.

Overall, we obtained responses from at least one committee member in the vast majority of participating schools in both years (Table 3.2). The respondents for whom we successfully determined school identifiers came from 191 schools in both years. It is possible that the respondents for whom we did not have sufficient information for school identification (15 respondents in 2008–2009 and two respondents in 2009–2010) might have included CC members in the other seven SPBP participating schools.

**Analysis and Reporting.** To adjust for potential differences due to nonresponse, we created weights so that our responding sample would be representative of the population of CC members in the 198 participating schools in 2008–2009 and the 196 schools in 2009–2010. However, the weighted results differed very little from the unweighted results. We have therefore reported the unweighted data in this report.[6]

In addition, we compared responses of individuals from different role groups (e.g., administrators and UFT-represented committee members) and from schools with different characteristics (e.g., level, size, demographics, bonus receipt). To simplify the presentation, we do not report tests of statistical significance here. However, as a general rule, we do explicitly discuss only statistically significant differences (at $p < 0.05$) and use the term "significantly" to indicate such differ-

[6] The difference between weighted and unweighted results was, on average, less than 0.05. Weights were calculated as the product of sampling weight and nonresponse adjustment. Because all committee members were contacted for this survey, the sample weight is 1 for all cases. The nonresponse adjustment was calculated by dividing the expected number of respondents by the actual number of respondents. Because we found significant differences in survey results by role but not by such other factors as school level and size, the nonresponse adjustment was calculated by role. Because the actual numbers of respondents differed by survey question, a nonresponse adjustment was calculated for each question.

**Table 3.2**
**Compensation Committee Survey Respondents, by School**

| | Total Schools | | Schools with | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | In Population | With Identifiable Respondents[a] | Four Respondents | | Three Respondents | | Two Respondents | | One Respondent | | Average Respondents per School |
| Year | N | N | No. | % | No. | % | No. | % | No. | % | |
| 2008–2009 | 198 | 191 | 61 | 32 | 51[b] | 26 | 46 | 24 | 33 | 17 | 2.7 |
| 2009–2010 | 196 | 191 | 65 | 34 | 68 | 36 | 37 | 19 | 21[c] | 11 | 2.9 |

[a] 15 and 2 respondents did not provide sufficient information for us to identify their school in 2008–2009 and 2009–2010, respectively.

[b] One of these 51 schools officially reported to the NYCDOE that they had 3 members—so in this case, we obtained responses from all members. There was one other school that officially reported to the NYCDOE having only 3 members but we either did not receive responses from this school or they are included in the group of respondents who did not provide sufficient information for us to identify their school and thus count the total respondents. We are not entirely sure why these two schools did not obtain the requisite 4 members.

[c] One school officially reported that they had 3 members on the Compensation Committee in 2009–2010. Only one committee member responded to the survey.

ences (e.g., saying that administrators were significantly more likely than UFT committee members to report). Because the large number of comparisons might have increased the probability of finding significant differences due to chance, a small percentage of the significant differences will likely be due to chance rather than to stable or reproducible differences among the groups. Readers should therefore interpret the discussions of significant differences cautiously, especially when the magnitudes of the differences are small.

All survey instruments also contained a question at the end for respondents to include additional comments or thoughts about SPBP. Although these open-ended responses are by no means generalizable, we use them here as a secondary source of data and include quotes from the responses to help illustrate findings from our other sources of data.

Response percentages were calculated using the number of respondents who answered each survey question. Thus, the total number of respondents to a specific survey question may differ from the total number of survey respondents. We have simplified tables throughout the text to allow easy comparison of results by committee member role (i.e., administrators and UFT staff). Please see Appendix A for more details about the results.

### Teacher Survey

To test whether SPBP had an effect on teachers that might have mediated or been the source of any effects on student achievement, we surveyed a sample of teachers in the second year of our study from all schools initially selected to participate in SPBP and from eligible schools that were not selected.[7] The survey also provided information about participating teachers' attitudes and perceptions about the program.

**Teacher Sample Selection.** At each school, we randomly selected four teachers of grades and subjects in which students take state tests and three teachers of untested grades and subjects. We believed this

[7] The teacher survey sample did not include a District 75 special education school because the student population and likely organization and instructional practices in this school were very different from those in the other schools in the sample.

stratification was important to ensure that we captured perspectives of teachers who may have felt different levels of pressure related to the Progress Report and bonus systems (i.e., unlike untested teachers, state test results for students in tested teacher classrooms directly affect schoolwide measures determining bonus receipt). Tested teachers were defined as teachers in grades 3–5, teachers of ELA and/or mathematics in grades 6–8, and high school teachers of any course culminating in a Regents Exam. All other teachers were classified as untested. For schools that had fewer than four tested teachers or fewer than three untested teachers, we selected all teachers from that category. In total, we sampled 2,692 teachers.

**Survey Instrument.** We developed two versions of the survey: One was the SPBP version, for teachers teaching in schools participating in the program during the 2009–2010 school year, and the other was the non-SPBP version, for teachers teaching in schools not participating in the program in 2009–2010. Both versions of the survey included a common set of questions covering instructional practices, data use, professional development, collaboration, school climate, accountability, and other areas. Additionally, the SPBP survey included questions measuring teachers' attitudes and perceptions concerning the program. As with the CC survey, we developed original items and adapted items from past surveys administered by RAND and Vanderbilt University to evaluate performance-pay programs and other interventions, including surveys for evaluations of the TEEG program; the Project on Incentive in Teaching experiment in Nashville, Tennessee; and the Round Rock Pilot Program on Team Incentives. Colleagues at RAND, as well as representatives of NYCDOE and UFT, reviewed the draft instruments. In addition, we pilot tested the survey with seven teachers, including four from schools participating in SPBP in 2009–2010 and three from other schools, and conducted cognitive interviews with each of the seven teachers to elicit feedback. Of these seven teachers, three taught in elementary schools, two taught in middle schools, and two taught in high schools. We used the feedback from the reviews and cognitive interviews to revise the survey.

**Survey Administration.** The survey was administered between March 15, 2010 and July 2, 2010. Teachers received an introductory

email on March 15 with a link to our online version of the survey. The online survey was created by RAND's Multimode Interviewing Capability group and was based on the research team's original paper version. Nonresponders were sent a series of reminders, including two follow-up emails, followed by a mailed letter including the URL of the survey and a password for the individual respondent, and finally a hardcopy version of the survey with a postage-paid return envelope sent via Federal Express. NCYDOE and UFT leaders also sent out email messages encouraging participation. RAND administered the survey and follow-up emails and provided technical assistance to survey respondents through a toll-free telephone hotline. RAND's Survey Research Group handled the mailing of hardcopy versions of the survey. After completing their surveys, all respondents received a $25 Amazon.com gift card.

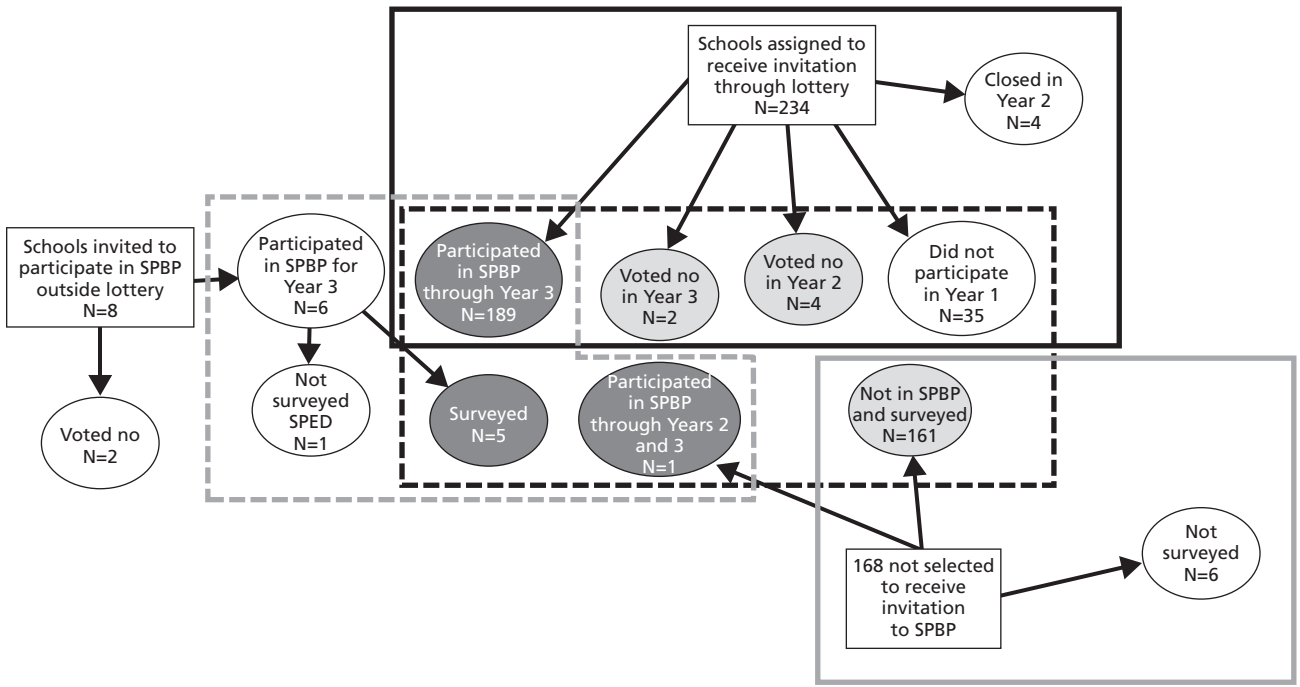
**Populations and Response Rates.** The teacher surveys supported two distinct sets of analyses, and we relied on slightly different groups of school populations for these two analyses. The first, presented in Chapters Four, Five, and Six, describes the experiences of teachers from the 196 schools that participated in SPBP during the 2009–2010 school year. These analyses relied on items that were common to both survey instruments and on those included only on the SPBP form. The second set, presented in Chapter Eight, estimates the program's effects on teacher behaviors, perceptions, and attitudes. The population of interest here was teachers in the 402 schools (234 assigned to SPBP and 168 assigned to control) that participated in the lottery. Because this analysis compares SPBP with control schools, it uses only items common to both survey instruments. There is overlap between the populations for the two sets of analyses, but the sample of SPBP-assigned and SPBP-participating schools are not the same because, as shown in Figure 2.1, only a subset of schools invited to participate in the program did participate, and only a small set of schools were invited to participate in the program outside the lottery process. Moreover, not every school in either group was included in the sampling frame for either analysis.

Figure 3.2 shows these school populations and samples for the teacher survey analyses to clarify the overlap or lack of overlap of these

groups. The 234 schools at the top, inside the rectangle outlined in a solid black line, represent the schools assigned through the lottery to receive an invitation to participate in SPBP. This is the population of schools for the SPBP effects analysis presented in Chapter Eight. The 168 schools at the bottom, inside the rectangle outlined in a solid gray line, represent schools that the lottery determined would not receive an invitation to participate in the program. This is the population of control schools for the SPBP effects analysis in Chapter Eight. Schools within the rectangle outlined by the dashed black line are schools that received the survey. These include 230 of the 234 SPBP schools, 162 of the 168 control schools, and 5 schools invited to participate in SPBP outside the lottery process. Schools within this rectangle in light gray circles did not participate in SPBP in 2010 and received the non-SPBP survey form.

Even though some schools were assigned to SPBP and would be part of the SPBP-assigned group in analyses, they received the non-SPBP form because they did not participate in SPBP in 2010. This does not affect the SPBP effect analysis in Chapter Eight because, as noted previously, the items used in that study were common across forms. The 196 schools in the rectangle outlined with the gray dashed line are those participating in SPBP in 2010. Of these, 195 received surveys and, since they were participating in SPBP, they received the SPBP form and are represented by dark gray circles. The weighted responses from these teachers are used in Chapters Four, Five, and Six to describe the experiences of SPBP teachers. Table 3.3 summarizes the survey response rates for the teachers surveyed for the SPBP effects analysis in Chapter Eight. Of the 390 schools included in the survey sample, the school-level response rate is roughly 100 percent for schools selected for SPBP and 99 percent for the control schools. The sample excluded four SPBP schools that closed, two SPBP schools, and six control schools that were mistakenly left off the list. The overall teacher-level response rate was 57 percent. The teacher-level response rates were 59 percent and 55 percent for the SPBP and control groups, respectively. To account for possible differential nonresponse between SPBP and other teachers, we controlled for available information on teachers and schools in our tests of program effects on survey responses.

**Figure 3.2**
**SPBP Teacher Survey Populations and Samples Used in the Analyses**




**RAND** *MG1114-3.2*

**Table 3.3**
**Overall Teacher Response Rates and Teacher Levels for Schools in the Lottery Process**

| | | | Teacher | | | |
|---|---|---|---|---|---|---|
| | School | | Treatment | | Control | |
| | Treatment | Control | Tested | Untested | Tested | Untested |
| Total number | | | | | | |
| Fielded | 228 | 162 | 896 | 670 | 637 | 475 |
| Responded | 227 | 160 | 544 | 379 | 351 | 258 |
| Percent | | | 61 | 57 | 55 | 54 |
| Percentage responding | 100 | 99 | 59 | | 55 | |

For the analyses of teacher experiences in the 196 schools participating in SPBP in 2009–2010 (findings presented in Chapters Four, Five, and Six), we received teacher responses from 100 percent of the 195 schools surveyed. One special education school was excluded because of its unique student population. For this sample of schools and teachers, we received responses from 817 teachers, representing 60 percent of the sample. Survey responses were weighted to account for teacher differential nonresponse across observed teacher and school characteristics, including whether or not the teacher taught a tested grade; school level or type (e.g., elementary, middle, K–8 or high schools); school-level aggregates of student characteristics, including race and/or ethnicity, free and reduced price meal status, ELL status, and special education status; and ELA or mathematics achievement.

It is worth noting that there may be some overlap in the survey samples of teachers and CC members in 2010; however, we believe that number to be quite small. On the teacher survey, 7 percent of teachers reported serving on the CC in 2009–2010. We do not know how many of them also filled out the CC survey. Yet, even if it were 100 percent, very few individuals would have filled out both surveys.

**Analysis and Reporting.** We conducted two sets of analyses on the teacher survey results. The goal of the first was to examine how

SPBP affected teachers (referred to as the effect-of-treatment analysis). A simple estimate of that effect is the difference in reported attitudes and behaviors for teachers in participating schools and those in eligible schools not participating. Ideally this approach would allow us to infer how much teachers' reported practices and attitudes would differ if their schools had or had not participated in SPBP.

However, as discussed in greater detail in Chapter Two, our analysis faced two challenges: schools' ability to select whether to participate in SPBP and the fact that some schools joined SPBP without being randomly assigned. Schools that did not follow their lottery assignments might differ systematically from other schools in the sample in ways that are related to measures of teacher attitudes and behaviors. Rather than use a simple analytical approach and risk unknown errors, we implemented a common alternative approach used for estimating treatment effects in randomized control trials, the "intent-to-treat" (ITT) analysis (Shadish, Cook, and Campbell, 2001). An ITT compares the average outcomes of schools assigned by lottery to SPBP, regardless of whether the school participated in the program in a particular school year, to the average outcomes for schools that were not selected by the lottery process to SPBP, regardless of whether the school participated in SPBP in a particular school year.

In total, teachers in SPBP and control schools completed over 80 common questions. We created 14 scales out of those common questions and conducted the effect of treatment analysis on the scale scores and 23 individual items that did not group into any scales.

For the effect-of-treatment analyses, we estimated the effects of participating in SPBP using 14 scales and 23 items.[8] We conducted a two-level mixed-effects hierarchical modeling analysis on the 14 scale scores and ordered logistic or logistic regression analysis, with robust standard errors, on the 23 items. Given the increased likelihood of finding a spurious significant treatment effect when the true effects are all null with multiple statistical tests, we used the Benjamini and Hochberg (1985) method to adjust for multiple comparisons. All models included an indicator for SPBP participation and controlled

[8] See Chapter Eight for the details.

for tested group, school grade level, and school-level demographic variables. School demographic variables included enrollment, percentage of ELL, percentage of students eligible for an individual education program, percentage of black and Hispanic students, and percentage of economically disadvantaged students. The model included these variables to account for potential differences between schools assigned to SPBP or and other schools that are due to chance variation from randomization or survey nonresponse. Details of the model specification are in Chapter Eight and Appendix E. The results of these analyses are presented in Chapter Eight. Appendix F contains more-detailed results of the effect from the treatment analysis.

The goal of the second set of analyses was to understand the responses of teachers in SPBP schools to the program. For this, the SPBP-participant-only analysis, we analyzed teachers' responses to the items that were unique to teachers in SPBP schools. We created weights so that our responding sample would be representative of the population of all classroom teachers in the SPBP schools. These results are presented in Chapters Four, Five, and Six. More detailed results are in Appendix B.

### Student Achievement Data and Analysis

The purpose of the student achievement analysis was to determine whether a school's SPBP participation meant that students enrolled in the school would have higher achievement scores than they would had the school not participated. We compared the test scores of students enrolled in schools invited to participate in SPBP with those for a comparable set of students and schools eligible for but not invited to participate in the program. The lottery selection process Roland Fryer implemented for NYCDOE helped ensure that the characteristics of both sets of schools were comparable. Further details of the data and analysis are provided in Chapter Seven.

**Data.** The analysis of how SPBP affected Progress Report scores and student test results required school-level and student-level data, including student characteristics and test scores. The former relied on the school-level data, while latter relied on both sets of data to identify and analyze test results for the students enrolled during the school

year just prior to SPBP and the three years of the program (i.e., the 2006–2007, 2007–2008, 2008–2009, and/or 2009–2010 school years) in schools entered into the SPBP lottery in fall 2007.

***School-Level Data.*** NYCDOE provided a series of files with data on school characteristics for the four school years of interest. These files contained demographic information on the students enrolled in each school, including student attendance rates, total student enrollment, and disciplinary incidents. The files also contained adequate yearly progress (AYP) status and ratings the school received under the State of New York's accountability program under No Child Left Behind (NCLB). To study the effects of SPBP on the Progress Reports, we downloaded all the relevant Progress Report scores and grades and the component scores and grades from NYCDOE website.[9] We obtained the data for the three program years and for the 2006–2007 year to uses for a baseline comparison of the SPBP assigned and control schools.

***Student-Level Data.*** In addition to the school-level data, analysis of student test scores required additional data on student characteristics and student test scores. NYCDOE provided two types of student-level data files: student biographical information and student achievement test scores. The student biographical files contained information on the universe of students enrolled in New York City public schools by school year and the school and grade in which a student was registered by semester. The biographical files also contained student background information, including gender, race and/or ethnicity, and participation in special education services and ELL programs, and eligibility for free or reduced priced school lunches. Both the biographical and achievement data files include a unique, longitudinally consistent identifier for each student. This meant that records on the same student could be linked across multiple schools and times and to records contained in separate data files.

---

[9] We downloaded the official results for elementary, middle, high, and transfer schools for the 2006–2007, 2007–2008, and 2008–2009 school years from the "Prior Year Progress" page on the NYCDOE website on February 26, 2011, and results for the 2009–2010 school year from "Progress Report" page on the same date. The school demographic information on the latter spreadsheet was incomplete and appeared to be preliminary. Hence, our results may not reflect the calculations in the final report.

The student test score files contained results for the universe of students enrolled in elementary, middle, and K–8 schools. NYCDOE also provided a similar set of files for high school students. For grades 3 through 8, the student files contained scores from the State of New York's mathematics and ELA assessments administered during the year just prior to SPBP and during the three years of the program. The high school student test files contained scores from the Regents Exams, Regents Competency Tests, Regents Portfolio Assessment, and approved alternative examinations for the year just prior to SPBP and the first two years of the program. Because of the timing of the release of the high school data for year 3, we were unable to include analyses of high school student achievement for that year (2009–2010).

**Basic Analytic Approach.** Similar to our approach for the teacher surveys, we used an ITT analysis to estimate the SPBP effect on student achievement. Again, a simple estimate of that effect is the difference between the average test score for students enrolled in participating schools and that for students enrolled in eligible schools not participating. Ideally, this approach would allow us to infer how much average student's test scores would differ, depending on whether their schools did or did not participate in SPBP. However, the various adjustments made to lottery assignments posed potential challenges to such an analysis. Rather than risk unknown errors, we implemented the ITT approach, comparing the average Progress Report scores and test scores of students enrolled in schools that were assigned by lottery to SPBP, regardless of whether the school participated in the program in a particular school year, to the average outcomes for schools that the lottery process did not designate for SPBP, regardless of whether the school participated in SPBP in a particular school year. Additional details on the data and the statistical methods used to estimate the effects are in Chapter Seven.

### School Case-Study Visits

To gain a more in-depth understanding of SPBP and staff experiences with the three years of the program, we conducted site visits to 14 schools, including seven in spring 2009 and seven in spring 2010. In these visits, researchers examined how individual schools implemented

the program (by asking questions about the current year and retrospective questions about previous program years) and inquired about staff perceptions of how the program was working and affecting them.

**Sample Selection.** In each study year, we selected seven schools that varied in basic school and student characteristics (school level, student enrollment, borough, student achievement, and student demographics) and the schools' experiences with SPBP. More specifically, the primary selection variables included

- *School level*. In the first study year, our sampling frame included three elementary, two middle, and two high schools. In the second year, we included four elementary, one middle, and two high schools. Given our small sample size, we excluded K–8 schools and schools with nonstandard grade configurations.
- *Receipt of bonus*. In the first study year, we sought to include schools that received the full bonus, partial bonus, and no bonus in year 1. (Note that, in the process of recruitment, we ended up with slightly fewer schools that won the full bonus than we had initially targeted.) In the second study year, we were interested in how patterns over time in whether a school won the bonus affected implementation in the school. For example, if a school lost the bonus two years in a row, did teachers in the school emphasize the program less? We therefore attempted to select schools that had a variety of experiences in terms of winning or not winning the bonus over time.
- *Previous year's distribution plan*. We analyzed data from NYCDOE to identify three categories of the previous year's plans: no differentiation, differentiation across job titles but not within job titles, and bonus differentiation within job titles. In selecting schools with bonuses that differentiated within job titles, we purposefully chose schools that exhibited the highest level of differentiation, paying particular attention to the number of bonus levels, the number of staff members at each level, and the extent to which staff within a job title received different bonus amounts. This strategy allowed us to obtain detailed information on schools that used highly differentiated distribution models and learn more

about the criteria schools used to decide to award certain staff members more or less of the bonus.

- *Participation in the program*. In the first study year, our sample included one of the three schools that had participated in year 1 and then dropped out of the program in year 2. In the second study year, our sample included only schools that were still participating in the program after three years.

Each year, after selecting an initial sample of schools using a combination of purposeful sampling and randomization techniques, we examined how representative this sample was in terms of a set of Secondary Selection Variables: Student Enrollment, Borough (Manhattan, Brooklyn, Bronx, Queens, and Staten Island), and 2007–2008 Progress Report Grade (A, B, C, D, or F). For each of the seven schools we initially selected each year, we also identified a backup school that mirrored the key characteristics (for a total of 14 schools) and that could have been recruited if the primary school had declined to participate.

In the first study year, although we initially contacted 14 schools to participate in the case studies, refusals and lack of response led us to contact an additional four schools, for a total of 18 schools. We finally recruited the seven schools for our case studies. In the second year, we once again ended up contacting 18 schools before recruiting the seven schools to participate in the case studies.[10] Table 3.4 summarizes the case-study samples by year according to the sampling criteria.

Tables 3.5 and 3.6 show how our case-study samples compared with the total population of participating schools. We compared the

[10] Although the refusal rate was quite high, we are not concerned that the sample we ended up with either year was biased in any particular direction. In the first study year, delays in starting the evaluation left very little time between the beginning of recruiting and the time we needed to begin site visits. We were not surprised that so many schools were reluctant to have us visit when we were giving them only a few weeks advance notice. Although at least one school we targeted because of its highly differentiated distribution plan from the first program year declined our request, the others did not indicate any consistent reason for their refusal. In some cases, we simply never heard back after making multiple calls. In other cases, the person contacted indicated that the timing at the end of the year was too difficult for a visit.

**Table 3.4**
**Summary of Case Study School Characteristics, by Selection Criteria and Year**

| | Case Study Schools | |
|---|---|---|
| **Variable** | **2009** | **2010** |
| School level | | |
| Elementary | 3 | 4 |
| Middle | 2 | 1 |
| High | 2 | 2 |
| Receipt of bonus | | |
| Full bonus in 2007–2008 | 2 | N/A |
| Partial bonus in 2007–2008 | 1 | N/A |
| No bonus in 2007–2008 | 4 | N/A |
| Full bonus in 2007–2008; full bonus in 2008–09 | N/A | 3 |
| Full bonus in 2007–2008; no bonus in 2008–09 | N/A | 1 |
| Partial bonus in 2007–2008; full bonus in 2008–09 | N/A | 2 |
| No bonus in 2007–2008; full bonus in 2008–09 | N/A | 1 |
| Previous year's distribution plan | | |
| No bonus differentiation in prior year | 2 | 1 |
| Differentiation across but not within job titles in prior year | 1 | 0 |
| Differentiation within job titles in prior year | 4 | 6 |
| SPBP participation in year of visit | | |
| Participated in program in year of visit | 6 | 7 |
| Did not participate in program in year of visit | 1 | 0 |

year 2 case-study schools to all schools participating in year 2 (even though one of our case-study schools dropped out of the program in year 2) and compared year 3 case-study schools to all schools participating in year 3. Given our small samples and the choice to oversample schools with highly differentiated bonus distribution models, we neither aimed nor expected to obtain a strictly representative sample of the total population of participating schools. Nevertheless, Tables 3.5 and 3.6 demonstrate that our samples did capture variations in the characteristics and experiences of SPBP schools.

**Table 3.5**
**School Characteristics: Comparison of All 2008–2009 Participating Schools with Spring 2009 Case Study Schools**

| | 2008–2009 Participating Schools (N=198) | | Spring 2009 Case Study Schools (N=7) | |
|---|---|---|---|---|
| **Variable** | **Number** | **Percent** | **Number** | **Percent** |
| School level | | | | |
| Elementary | 92 | 46 | 3 | 43 |
| Middle | 39 | 20 | 2 | 29 |
| K–8 | 21 | 11 | 0 | 0 |
| High | 34 | 17 | 2 | 29 |
| MS/HS | 9 | 5 | 0 | 0 |
| Transfer | 3 | 2 | 0 | 0 |
| Bonus result (2007–2008) | | | | |
| Schools that | | | | |
| Received the full bonus | 92 | 46 | 2 | 29 |
| Received the partial bonus | 30 | 15 | 1 | 14 |
| Did not receive a bonus | 76 | 38 | 4 | 57 |
| Bonus distribution plan (2007–2008)[a] | | | | |
| Schools with | | | | |
| No bonus differentiation | 31 | 16 | 2 | 29 |
| Bonus differentiation across job titles but not within job titles | 44 | 22 | 1 | 14 |
| Bonus differentiation within job titles | 123 | 62 | 4 | 57 |
| Borough schools | | | | |
| The Bronx | 79 | 40 | 4 | 57 |
| Brooklyn | 64 | 32 | 1 | 14 |
| Manhattan | 42 | 21 | 1 | 14 |
| Queens | 11 | 6 | 1 | 14 |
| Staten Island | 2 | 1 | 0 | 0 |

**Table 3.5—Continued**

| Variable | 2008–2009 Participating Schools (N=198) | | Spring 2009 Case Study Schools (N=7) | |
|---|---|---|---|---|
| | Number | Percent | Number | Percent |
| Student characteristics, 2007–2008 averages[a] | | | | |
| Enrollment | 642 | | 592 | |
| Black or Hispanic | | 96 | | 96 |
| Free and reduced–price lunch | | 83 | | 83 |
| ELL | | 19 | | 16 |
| Special education | | 19 | | 22 |
| Progress report grade (2007–2008)[b] | | | | |
| A | 63 | 34 | 1 | 14 |
| B | 83 | 44 | 3 | 43 |
| C | 25 | 13 | 2 | 29 |
| D | 14 | 7 | 0 | 0 |
| F | 3 | 2 | 0 | 0 |

[a] Demographic data were missing for six schools not in our sample. The averages for these population variables are based on n=192.

[b] One case study school did not have a progress report grade. In the population, ten schools did not have a progress report. The averages for these population variables are therefore based on n=188.

A team of four researchers conducted the visits both years, with two researchers visiting each school. Researchers interviewed respondents one on one (with a few exceptions) and used semistructured interview protocols that had been developed for each respondent group. Interviews ranged from 20 minutes to an hour depending on the type of respondent and the individual's knowledge of the program. We asked permission to audio record the interviews, and most interviewees agreed to be recorded. All participants were promised anonymity.

**Analysis.** Following the visits, case-study research teams analyzed interview notes and developed detailed analytic memoranda for each school visited. The memoranda analyzed interviewee reports along a set of dimensions, including year 1 and 2 implementation (level of

**Table 3.6**
**School Characteristics: Comparison of All Schools 2009–2010 Participating with Spring 2010 Case Study Schools**

| | 2009–2010 Participating Schools (N = 196) | | Spring 2010 Case Study Schools (N = 7) | |
|---|---|---|---|---|
| **Variable** | **Number** | **Percent** | **Number** | **Percent** |
| School level | | | | |
| Elementary | 91 | 47 | 4 | 57 |
| Middle | 39 | 20 | 1 | 14 |
| K–8 | 21 | 11 | 0 | 0 |
| High | 34 | 17 | 2 | 29 |
| MS/HS | 8 | 4 | 0 | 0 |
| Transfer | 3 | 2 | 0 | 0 |
| Bonus result (2007–2008 and 2008–2009) | | | | |
| Schools receiving | | | | |
| Full bonus in years 1 and 2 | 62 | 32 | 3 | 43 |
| Full bonus in year 1 and partial bonus in year 2 | 4 | 2 | 0 | 0 |
| Full bonus in year 1 and no bonus in year 2 | 26 | 13 | 1 | 14 |
| Partial bonus in year 1 and full bonus in year 2 | 25 | 13 | 2 | 29 |
| Partial bonus in years 1 and 2 | 2 | 1 | 0 | 0 |
| Partial bonus in year 1 and no bonus in year 2 | 3 | 2 | 0 | 0 |
| No bonus in year 1 and full bonus in year 2 | 70 | 36 | 1 | 14 |
| No bonus in year 1 and partial bonus in year 2 | 2 | 1 | 0 | 0 |
| No bonus in years 1 and 2 | 3 | 2 | 0 | 0 |
| Bonus distribution plan (2008–2009) | | | | |
| Schools with | | | | |
| No bonus differentiation | 30 | 15 | 1 | 14 |
| Bonus differentiation across job titles but not within job titles | 43 | 22 | 0 | 0 |
| Bonus differentiation within job titles | 123 | 63 | 6 | 86 |

**Table 3.6—Continued**

| Variable | 2009–2010 Participating Schools (N = 196) | | Spring 2010 Case Study Schools (N = 7) | |
|---|---|---|---|---|
| | Number | Percent | Number | Percent |
| Borough | | | | |
| The Bronx | 79 | 40 | 5 | 71 |
| Brooklyn | 63 | 32 | 2 | 29 |
| Manhattan | 41 | 21 | 0 | 0 |
| Queens | 11 | 6 | 0 | 0 |
| Staten Island | 2 | 1 | 0 | 0 |
| Student characteristics, 2007–2008 averages [a] | | | | |
| Enrollment | 646 | | 623 | |
| Black or Hispanic | | 96 | | 97 |
| Free– and reduced–price lunch | | 84 | | 83 |
| ELL | | 19 | | 17 |
| Special education | | 19 | | 17 |
| Progress report grade (2008–2009)[b] | | | | |
| A | 147 | 79 | 5 | 71 |
| B | 21 | 11 | 1 | 14 |
| C | 9 | 5 | 1 | 14 |
| D | 8 | 4 | 0 | 0 |
| F | 1 | 1 | 0 | |

[a] Demographic data were missing for six schools not in our sample. The averages for these population variables are based on n=190.

[b] Progress report data were missing for ten schools. The percentages for these variables are therefore based on n=186.

support for participation, experience of CC members, distribution plan, receipt of bonus), communication and technical assistance; data and tools, communication to stakeholders; perceived effects (effects on staff interactions, effects on recruitment and retention, individual affective and behavioral effects), general attitudes and opinions about SPBP, and advice for improving the program. The memoranda detailed the number and types of individuals reporting various opinions and

included verbatim excerpts from the recorded interviews as supporting evidence. (Table 3.7 tallies the types of participants overall.) Each year, researchers then analyzed the evidence across schools to develop cross-case themes and findings.

### Leader Interviews

To learn more about the history of SPBP, its goals and the mechanisms intended to achieve them, and the perceptions about early successes and challenges, we conducted telephone interviews in the first study year with representatives from all organizations sponsoring or participating in the program. This included leaders from NYCDOE, UFT, the Broad Foundation, and the Fund for Public Schools. We conducted a total of 13 interviews with 11 individuals, audiotaping and transcribing all interviews. We then analyzed the transcripts to understand and articulate the history and design of the program and its theory of action.

In our second study year, we conducted four follow-up telephone interviews with leaders from NYCDOE, UFT, and CSA to learn about SPBP year 3, potential changes to the program, perceptions about SPBP's effectiveness, lessons learned, and thoughts about its future.

### Document Review

To further examine the history, goals, and early implementation of the program, we collected documents from a number of sources, including NYCDOE and UFT staff and websites, local newspapers, and online news sources. These documents provided useful information about the program's design and how the program was described to participating schools and the public and the resulting response.

### Meeting Observations

In conjunction with the administration of CC surveys, we observed all three NYCDOE and UFT-sponsored spring 2009 meetings for CC members and gathered copies of the slides presented. This provided useful information on how leaders were communicating about the program.

**Table 3.7**
**Case Study Interviewees, by School and Respondent Group**

| | Spring 2009 Case Studies | | | | | | | Spring 2010 Case Studies | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Respondent Type** | **S1** | **S2** | **S3** | **S4** | **S5** | **S6** | **S7** | **S8** | **S9** | **S10** | **S11** | **S12** | **S13** | **S14** | **Total** |
| Principals | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 14 |
| Assistant principals | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 10 |
| Classroom teachers | 5 | 6 | 6 | 5 | 5 | 5 | 5 | 6 | 4 | 6 | 5 | 5 | 4 | 5 | 72 |
| Non–classroom teachers[a] | 2 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 4 | 1 | 2 | 0 | 0 | 0 | 12 |
| Clinical and student support staff[b] | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 11 |
| Secretaries | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 0 | 0 | 1 | 0 | 0 | 7 |
| Paraprofessionals | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 0 | 0 | 5 |
| Total | 10 | 10 | 9 | 9 | 9 | 9 | 10 | 9 | 10 | 10 | 11 | 10 | 7 | 8 | 131 |

[a] These teachers include data specialists, reading and math coaches, and other specialists.

[b] These personnel includes guidance counselors, social workers, and others.

### Administrative Data and Bonus Distribution Plans

To understand how CCs planned to distribute bonuses among staff members, assuming their schools won the bonus,[11] we collected and analyzed the information CCs reported to NYCDOE in spring 2008, 2009, and 2010. For each year, the data included the amount allocated to each UFT-represented staff member and the members' position title. The data also contained the school name and unique school identification.

For our analyses of how SPBP participation affected student achievement, NYCDOE provided a series of files that contained information on all schools meeting the criteria to be eligible for SPBP and the outcomes of the lotteries it conducted to select the schools it would invite to participate in the program. The data also included information on the staff votes on participation each year and various anomalies in the SPBP lottery and invitation processes.

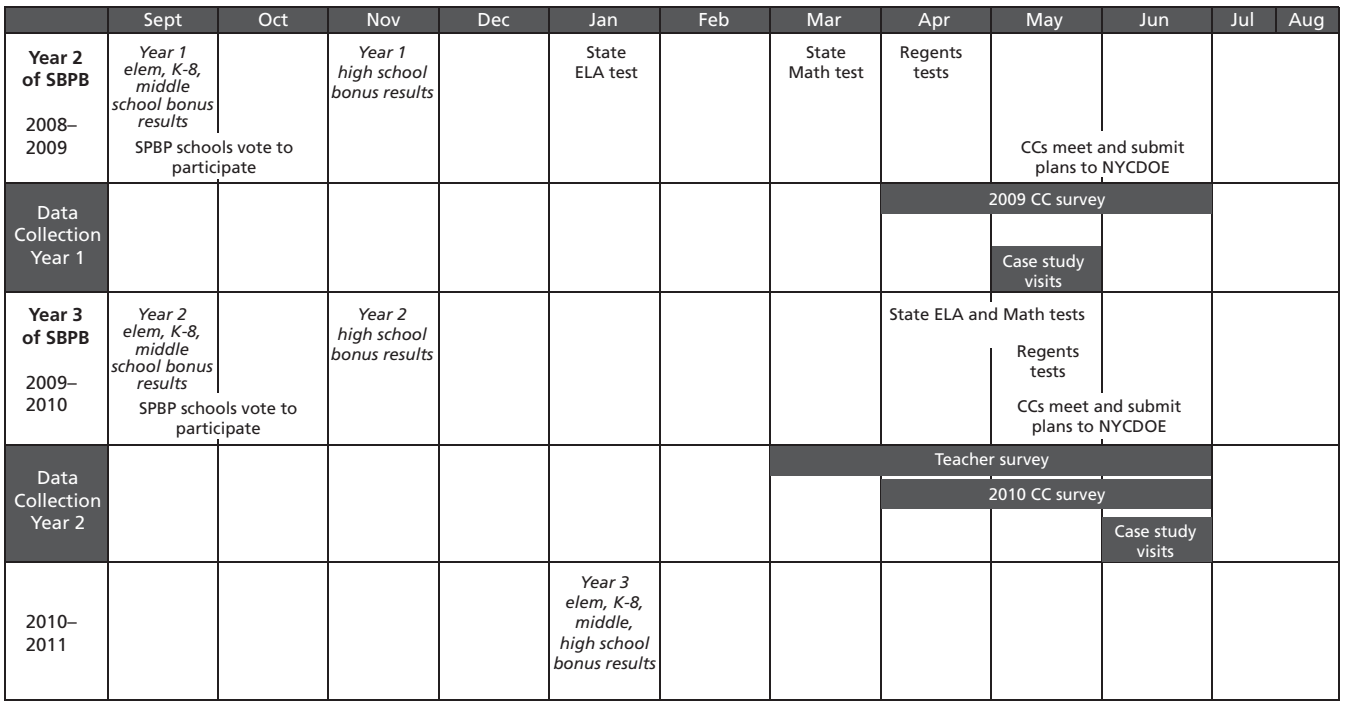
### Timeline of School-Level Data Collection Activities

As Figure 3.3, illustrates, our data collection activities in schools coincided with SPBP years 2 and 3. In the first year of our study (the 2008–2009 school year, SPBP year 2), the CC survey was administered, and case-study visits occurred many months after year 1 bonus results were announced but before year 2 results were announced. Both these data-collection activities occurred just after state testing in elementary, middle, and K–8 schools and right around the time of Regents testing in high schools but before the announcement of those results. In the second study year (the 2009–2010 school year, SPBP year 3), the teacher and CC surveys were administered and case-study visits occurred many months after year 2 bonus results were announced but before year 3 results were announced. These data collection activities coincided with the administration of state tests but prior to the announcement of these results.

Given this timing, our survey and interview questions in some cases referred to the previous program year (e.g., Were you satisfied with the bonus amount you received from last year's bonus?) and in

[11] All schools were required to develop such plans before the end of each school year.

**Figure 3.3**
**Timeline of Data Collection Activities Relative to SPBP Program Timeline**




RAND *MG1114-3.3*

others focused on the current program year (e.g., If your school wins the bonus this year, how will the committee divide up the funding among staff?). As illustrated in the figure, our data collection activities did not capture staff responses to the results of year 3.

The distribution plans the CCs created were entered into the Galaxy system at the end of the school year, generally after respondents took the surveys. The plans for distributing awards would be known to school staff only in the next school year, after the bonuses were awarded.

## Study Limitations

Because of limited resources and time constraints, we were unable to survey teachers in SPBP and control schools in the first year of the evaluation and thus have only one year of teacher data. However the research team did collect two years of data from the population of CC members (a maximum of four individuals per school) and some additional staff members in the case-study schools. Although year 1 data may not represent the views of all staff members in SPBP schools and do not provide insights into non-SPBP schools, they do provide important descriptive information from a variety of administrators, teachers, and other UFT staff members in almost all SPBP schools. We surveyed a random sample of teachers in all SPBP and control schools in the second year, which provided more-representative estimates of attitudes, activities, behaviors, and perceived effects within and across both sets of schools. However, these survey data are not representative of nonteaching staff in the schools. Again, resource constraints would not allow for a broader survey of all school staff, and opinions from nonteachers thus come only from the case studies and, in some cases, the CC surveys.

Further, our response rate of 57 percent is another limitation on these data. We controlled for possible differential response rates on observed characteristics by including such variables in the models for testing SPBP effects and by nonresponse weighting in our analysis of the teacher populations in SPBP schools. Nonresponse weighting

cannot account for differences among respondents and other teachers on factors we did not observe, and given the relatively low response rate, such differences could result in sizeable biases. We cannot assess the potential for such biases. We know the respondents and other teachers were generally similar on observed data, but that does necessarily imply similarity on the unobserved factors. Similarly, including observed variables in the models testing for SPBP effects can only control differences among the groups on observable characteristics. Differences on unobserved variables could still lead to potential biases in our estimated SPBP effects. Since we observed almost no group differences, biases due to differential response would need to be masking true effects, and the null findings alone should not be taken as evidence that there were no SPBP effects. However, as discussed in Chapters Four through Seven, the null findings generally are consistent with other data from the school staff and our student outcome analyses.

Also, across both sets of surveys, we do not have pre-post measures of attitudes and practices, and our data measure impressions of change post hoc. Such impressions are subject to the halo effect attributable to receiving or not receiving a bonus. We also acknowledge that surveys may not always accurately measure some of the constructs of interest in this study, such as instructional practices and motivation, but these limitations of surveys should not differ between SPBP and other teachers.

Another limitation here is that our analysis of how SPBP affected student outcomes relies exclusively on state test scores and does not include other individual student-level outcomes, such as attendance and discipline. Test scores may capture only a portion of student achievement and may not measure all valuable outcomes. However, other factors, such as attendance, discipline, and graduation rates, are included in Progress Report scores. Our model for program effects tests for these in aggregate on the reports.

It is worth noting one final caveat about interpreting survey and interview responses from the study's final year of data collection. Surveys and school visits in year 3 (2009–2010) took place the year after the vast majority of schools won the bonus in year 2. As noted in Chapter Two, in the same year, educators throughout the state were told

that the state would be changing cut scores on the state standardized achievement tests for 2010–2011.[12] This context could have directly influenced responses to some of our questions (e.g., perceptions about attainability of performance targets and receiving bonuses).

## Summary

In summary, this evaluation used a combination of quantitative and qualitative methods to examine the following research questions:

1. How was the SPBP program implemented?
2. What were the intermediate outcomes of the program?
3. How did the program affect student performance?

To understand the implementation and perceived effects of the program in years 1 through 3, we collected and analyzed data from surveys of CC members and teachers in SPBP and control schools; interviews in 14 case-study schools; primary documents; administrative data from NYCDOE; and interviews with NYCDOE and UFT leaders and funders. To understand the effects of the program on school and staff practices, we analyzed survey data from teachers in SPBP and control schools. To understand the program effects on student outcomes and achievement, we analyzed the Progress Report scores and student test results and demographic data for participating and control schools from the 2006–2007 to 2009–2010 school years.

In the next three chapters, we examine how school-level staff implemented and responded to SPBP, including staff attitudes about the program and the bonus, the nature of communication about and staff understanding of the program, the CC process, distribution plans, and perceptions about effects. These three chapters examine important factors that may mediate the intended effects of SPBP.

[12] NYCDOE staff communicated with principals and networks throughout the year that there would be changes to the cut scores.

CHAPTER FOUR

# Implementation of the Schoolwide Performance Bonus Program: Attitudes About and Understanding of the Program

With any program, ultimate success is often influenced by the level of support for and understanding of the particular program and its main components among individuals targeted by or who are expected to implement the program. As discussed in Chapter Two, theory and research on pay-for-performance programs indicate that how individuals respond to bonus incentives and the extent to which the incentives provide motivational effects are likely to be influenced by the awareness and support for the program's purposes and components, the perceived fairness of the methods for awarding bonuses, the perceived value of the bonus, beliefs about whether individual efforts can improve student performance to meet program goals, the perceived attainability of the performance targets, understandings about how performance will be measured, and the criteria for receiving the bonus. Individuals may be less inclined to work toward the accountability targets if they lack this basic awareness, support, and understanding.

In this chapter, we examine attitudes and understandings that may affect responses to the bonus incentive and SPBP overall, starting with the reported level of support for the program, then attitudes about the bonus, and finally, the level of understanding of the program. In the process, we also answer key questions about the implementation of the program, including the nature of communication.

As we will illustrate, although staff conveyed high levels of support for the program and its continuation, they also voiced concerns

about particular features of the program. Attitudes about the bonus were also mixed. Although the majority of staff reported a strong desire to earn a bonus and although most winning the bonus were satisfied with the amount, many voiced concerns about the value and salience of the bonus. Finally, although most educators we interviewed and surveyed reported that many of the initial problems with communication were corrected over time, many misunderstandings about the program lingered. Collectively, these findings indicate some conditions that were supportive of SPBP and the intended motivational effects, and other conditions that were less conducive.

This chapter addresses the following research questions:

1. What were the levels of awareness of, support for, and attitudes about the program?
    a. What were the overall vote counts in schools?
    b. Did school staff support key components of the program?
    c. Did they support the criteria and measures used for determining bonuses?
        - Did they support the composition of the CCs?
        - Did they support the guidelines for staff eligibility?
    d. Did they think the program was fair?
2. What attitudes did staff members have about the financial bonuses?
    a. Did staff members have a desire to earn the bonus?
    b. To what extent were they satisfied with the amount of the bonus?
    c. How salient was the bonus? How did it rank among other possible motivating factors?
    d. What was the level of expectancy among staff? Did they feel that receiving a bonus was attainable?
3. What were the nature of communication and the level of understanding of the program?
    a. How did NYCDOE and UFT communicate with schools about the program?
    b. Did the schools perceive this information to be sufficient?

c. How well did staff members understand the program, including the Progress Report measures used to determine the bonus? Did that change over time?

Throughout this chapter (and the subsequent two), we draw on CC survey data and from visits to the 14 case-study schools over the two years of our study. We also present some results from items included in the surveys of SPBP teachers. Whenever relevant, we also present impressions from district and UFT leaders interviewed for our study. While much of these data focus on year 3, we also include some retrospective data about years 1 and 2. We present CC survey data primarily from the third and final year (2009–2010), making note when there are significant differences from data collected the prior year. Overall, these measures appeared to be quite stable over time, and we found very few differences in the responses of CC members in the first and second years of the study.

We also present comparisons of survey responses from different types of respondents. Although we found no statistically significant associations between teacher and CC member responses and student demographics (percentage of students who are ELL, black and Hispanic, or eligible for free or reduced-price lunch) or school size (with the exception of two items, in which the magnitude of difference was very small), we did find some significant differences between administrators and UFT-represented committee members (hereafter "UFT committee members"), respondents at different school levels (elementary, middle, high), and respondents receiving larger as opposed to smaller bonuses. We also present some comparisons of teachers from schools that won a full or partial SPBP bonus in year 2 and that did not win the bonus.[1] These comparisons should, however, be interpreted cautiously. The significant associations do not indicate causation. Differences between these groups could be a consequence of bonus results (e.g., teachers in nonbonus schools may have more-negative attitudes

[1] We do not include comparisons of CC members from bonus and nonbonus schools because the number of schools in the latter category was too small to allow detection of significant differences among CC members (ranging from one to four individuals per school). The larger number of teachers surveyed per school allowed greater statistical power.

or may have lacked understanding because they were disappointed with the outcome and therefore felt more negative about the program), a cause of bonus results (i.e., the negative attitudes or lower level of reported understanding of the program caused the school to not win the bonus), or neither (e.g., teachers in nonbonus schools would have been more negative about these things with or without the program and the bonus outcome). Given the data available for this evaluation, we cannot determine the nature of the relationships described previously or present them to clarify differences among schools.

## General Attitudes

This section describes staff awareness of and attitudes and opinions about SPBP. It provides vote count, survey, and case-study data indicating high levels of support for the program and its continuation and some concerns about particular features of the program.

### Our Data Provided Mixed Evidence of Staff Awareness of the Program: While Most Staff Knew That Their Schools Were Participating, SPBP Did Not Appear to Be Central to Staff Conversations in Many Schools

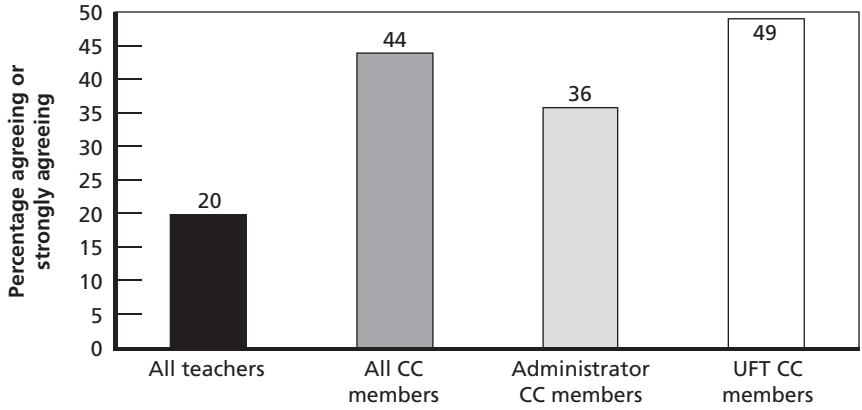
In case-study visits, the majority of interviewees were aware of the program and their schools' participation in it. In general, case-study respondents reported that SPBP was not on their minds and was rarely discussed in school. Some principals were also adamant that they did not mention the bonus and did not see it as a motivating tool. In other schools, principals who had championed differentiated bonus plans reported that they strategically avoided communication about the program to minimize potential problems among staff. Some fallout among staff did occur in one case-study school after it differentiated the distribution of bonuses in year 1; the principal then purposefully did not communicate about or explicitly discuss SPBP in staff meetings in year 2. The principal feared that focusing on the program would reignite some of the resentment that staff members who had received smaller shares of the bonus the previous year had felt. "I didn't

want to blow any more wind on the embers," he explained, "For me, I needed to let the fire die out and get everyone focused back on what is important, which is teaching and learning." Similarly, the principal at a spring 2010 case-study school purposefully avoided communication to minimize potential tension related to the differentiated plan. In fact, many educators in this school reported not knowing about their participation in year 3 (this particular school did not take a formal vote to participate).

In contrast, other principals appeared to invoke SPBP in meetings as one of many motivating tools, including talk about closure and other sanctions. One principal likened his communication about SPBP to a method actor trying to identify "whatever the motivation that can get you to act" and hoping that his staff could "find something in the menu here that resonates" for them.

As illustrated in Figure 4.1, the survey data indicate that CC members were much more likely than teachers to communicate with colleagues about SPBP. Only 20 percent of teachers said they frequently

**Figure 4.1**
**Percentage of Teachers and CC Members Agreeing or Strongly Agreeing with the Statement "I Frequently Talk with Colleagues in My School About the SPBP" (2010)**




NOTE: Total number of teacher respondents was 767. Total number of CC respondents was 531. The difference between administrator CC members and UFT CC members was statistically significant.

RAND *MG1114-4.1*

talked with colleagues about SPBP, compared to 44 percent of CC members (36 percent of administrator CC members and 49 percent of UFT CC members). Given their responsibilities on the committee related to administering the program, one would expect CC members to talk about the program more frequently than teachers. However, it is surprising that 56 percent of CC members reported infrequent to no talk. It is also not clear from this survey item whether the content of the talk was positive or negative.

### Staff Were Generally Supportive of Participating in the Program and of Its Continuation

Ninety-eight percent of all schools participating in year 1 voted to participate again in year 2. Among the schools for which vote counts were available, only a minority of UFT-represented staff appeared to have voted against participation. As Table 4.1 illustrates, in almost two-thirds of the schools that participated in year 2, more than 90 percent of the staff indicated that they wanted their schools to participate. The proportion of school-level "yes" votes among schools ranged from 55 to 100 percent, with an average of 90 percent overall (recall that it required a 55 percent yes vote to participate in the program).[2] The results for year 3 were very similar. Of the 182 schools that we have data for, more than 90 percent of staff voted to participate in almost two-thirds of the schools.

---

[2] NYCDOE did not systematically record year 1 vote counts for all schools invited to participate in the SPBP, which prevented us from using this information to determine the relative support in year 2 compared to year 1. Also, there appears to be some inconsistency in how votes may have been taken at schools, particularly in year 1. UFT and NYCDOE indicated to some schools, when asked, that staff not present at the time of the vote be counted as "no" votes. One leader explained that this was intended to encourage all staff to weigh in on the decision. However, this was not consistently enforced, and most schools appeared to submit vote counts to NYCDOE that included a vote of those present and did not count those absent as no votes (e.g., emails to NYCDOE would say, "Of the XX staff present, XX voted yes, and XX voted no"). Furthermore, in year 2, NYCDOE documented that one school had 20 staff not present at the time of the vote who were initially counted as "no" votes and meant the school would not participate by a vote of 16 to 26. However, the school initiated a revote which favored participation by a vote of 31 to 6. Because we cannot assess how widespread this practice may have been, it is not clear how precise the counts displayed in Table 4.1 are and the extent to which schools counted staff not present at the time of the vote as "no" votes.

**Table 4.1**
**School-Level UFT-Represented Staff Votes to Participate in Year 2 and Year 3 of SPBP**

| Percentage of Staff Voting Yes | Schools Year 2 | | Schools Year 3 | |
|---|---|---|---|---|
| | Number | Percent | Number | Percent |
| 55 to 60 | 5 | 3 | 1 | 1 |
| 61 to 70 | 14 | 7 | 16 | 9 |
| 71 to 80 | 13 | 7 | 16 | 9 |
| 81 to 90 | 37 | 19 | 32 | 18 |
| 91 to 95 | 46 | 24 | 24 | 13 |
| 96 to 100 | 78 | 40 | 93 | 51 |

NOTES: Year 2 data come from the 193 schools that reported vote tallies to the NYCDOE. Year 3 data come from the 182 schools that reported vote tallies to the NYCDOE.

Despite strong support for the program, some schools voted not to participate in SPBP during all three years of implementation. In year 1, 34 schools invited to participate in SPBP voted not to participate, two schools withdrew from the program after initially voting to participate, and one additional school had to drop out of the program because the school's CC could not reach consensus on a distribution plan. In year 2, an additional four schools declined to participate. In year 3 two schools did not to participate.

According to NYCDOE and UFT leaders, schools had one of two reasons for refusing to participate in year 1. One reason schools gave was that teachers at the school had a philosophical objection to a program they viewed as merit pay: "There were some schools that took a principled stand against [the plan] because they called it merit pay, and they thought it was a slippery slope," commented a UFT official. A second reason a few schools chose not to participate in round one of SPBP was related to school-based labor-management relations. SPBP requires cooperation among teachers and between teachers and the principal. "If the administration and the teachers did not play well together," said one UFT leader, "the prospects of the program being successful in that school were greatly reduced."

These same NYCDOE and UFT leaders also indicated that support for the program increased over time. Some of the schools that chose not to participate in SPBP in year 1 wanted to participate in years 2 and 3. In fact, many more eligible schools have expressed interest in participating. Said one NYCDOE official, "We have way more schools [interested in participating] than we have money." Echoed a UFT leader, "We get all the time from people, 'How can we get in the program?'" Another UFT leader concurred, "The schools in the bonus program are definitely happy. And other schools want access to it [the program]."

In most case-study schools in both years, one or two individuals interviewed disagreed with the program in principle or were concerned about certain aspects of the program. For example, one high school teacher was ideologically opposed to the program, noting that "it does look like some sort of merit-based pay [and] I don't believe it belongs here." Several respondents felt strongly that they should not receive bonuses for a job for which they are already compensated. "I don't believe in it [the SPBP]," explained an assistant principal at another high school, "It's like asking me, 'I'll give you a Hershey's bar if you do this.' It's my job. I'm supposed to do it. I don't need a carrot at the end of a stick to get me to do something I'm supposed to be doing already."

A small number of CC survey respondents also expressed a philosophical opposition to SPBP. For example, one elementary school teacher wrote:

> It is short sighted to think teachers change their practice because of some silly bonus program. To be honest, we don't ever think of it. We want students to achieve for them, not some token from the NYCDOE and UFT. This program, in my opinion is somewhat divisive and . . . should be rethought.

However, in both study years (years 2 and 3 of the program), the majority of case-study interviewees approved of their school's participation and wanted the program to continue. A minority of individuals expressed strong support for the program, noting in some cases that SPBP challenged them to push for better results or put them in a "collective mindset" to work for change. More often, this support did not

take the form of a wholesale endorsement of the underlying assumptions of SPBP (e.g., that the bonus will create more collaboration or improve behavior). Instead, the majority believed that staff worked hard and deserved recognition and that there was nothing to lose in participating in the program. As one elementary school teacher stated, "People felt that you do the work and you're working hard and someone is going to reward you for the hard work that you already do. . . why would you not want to be a part of that?" A teacher in another school similarly noted, "Sure, who wouldn't like to be rewarded and acknowledged in a concrete way?"

Interestingly, in a few schools, several staff members indicated that they were unaware of the requirement to vote in year 2. They had assumed that, once they voted to participate in the program in year 1, their schools would continue to participate in subsequent school years. In fact, in two schools visited in 2009, some school leaders who did not support SPBP attempted to stop the school's participation in the program being brought up for a vote, in the hopes that staff would not realize that the school had become eligible (for the vote in year 1) or was required to vote annually (for the vote in year 2). In one school visited in 2010, no formal vote occurred in year 3.

### Although More Than One-Half of All the Teachers and CC Members Believed SPBP Was Fair to All Schools, Some Staff—Particularly Teachers and UFT-Represented CC Members—Voiced Concerns About the Program Criteria, Timing, CC Requirements, and Staff Eligibility

As Table 4.2 illustrates, just over one-half of teachers and the majority of CC members believed SPBP was fair to all schools. In addition, many respondents expressed concerns about aspects of the program.

**Concerns About Criteria.** Many individuals questioned the metrics and the criteria chosen to evaluate the performance of schools. First, 84 percent of teachers and 74 percent of CC survey respondents felt the criteria for awarding the bonus relied too heavily on student test results. This was particularly true for UFT committee members, who were significantly more likely to voice this concern than were their administrative counterparts on the committee (78 percent to 68 per-

**Table 4.2**
**Percentage of Teachers and CC Members Agreeing or Strongly Agreeing with Statements About the Schoolwide Performance Bonus Program (2010)**

| | | Compensation Committee | | |
|---|---|---|---|---|
| **Statement** | **Teachers** | **All Members** | **Administrators** | **UFT Staff** |
| It would be helpful if staff knew at the start of the school year what the plans were for distributing the SPBP bonus award and how much additional pay an individual could earn | 84 | 77 | 77 | 77 |
| The criteria for a school to get a SPBP bonus award relies too heavily on student test results | 84 | 74 | **68** | **78** |
| Rewarding teachers based on test scores gains is problematic because it is hard to relate gains in student achievement to the work done by teachers | 82 | 73 | **57** | **84** |
| The UFT–represented employees should have more than half of the representatives on the Compensation Committee | 70 | 49 | **33** | **59** |
| The SPBP method for awarding bonuses (based on Progress Report targets) is fair to all participating schools | 54 | 78 | 79 | 77 |
| The SPBP program guidelines should be altered to ensure that bonus awards only go to staff who truly contribute to student success | 38 | 40 | **52** | **32** |
| Our school's Progress Report target is too high | NA | 43 | 43 | 44 |
| I would rather have an individual performance–pay program than the SPBP | NA | 26 | **46** | **14** |

NOTES: Statistically significant differences between Administrators and UFT staff are indicated in boldface ($p < 0.05$). Total numbers and percentage calculations for each item exclude those who responded "don't know." See Appendixes A and B for more detailed results on these items.

cent). Similarly, 82 percent of teachers and 73 percent of CC respondents agreed that "rewarding teachers based on test scores gains is problematic because it is hard to relate gains in student achievement to the work done by teachers." Once again, UFT committee members were significantly more likely than administrators to agree with this statement (84 percent to 57 percent).

Case-study respondents echoed these concerns. In one case-study school, many staff members believed the process was arbitrary and unfair and claimed that certain types of students were not properly counted. In other schools, staff expressed concerns about an overreliance on test scores as the basis for rewarding schools. On the other hand, one case-study principal felt strongly that the school environment survey should be weighted less in or removed entirely from the formula and that these results should be used for formative purposes of learning rather than for judging and rewarding a school.

In year 3, a handful of case-study interviewees voiced concerns about the lack of information from NYCDOE on new bonus criteria and targets for 2009–2010. One teacher on an elementary school CC explained, "I'd like to know what the bar is, because [someone at the UFT meeting] mentioned the bar would be higher. . . . If you have a goal, you need to know what to do to get there. You need to know the goals." A colleague echoed this concern: "I'd like to know what the targets are before when we go in. . . . without knowing specifically what the criteria is [sic] we have nothing to actually work for." Similarly, a teacher from another elementary school noted, "I would like to know the specific criteria, then it might change my goal." Others implied either a lack of awareness that criteria were potentially changing or disinterest. As one teacher explained, "I don't know if it would change really what I am doing . . . I guess it's good to know, but I don't think it would really change much."

The last sentiment fits well with the perceptions of NYCDOE and UFT leaders, who believed that "people don't really pay attention to the targets anyway" but instead strive for improvement and assume that, if students improved on the tests, they would qualify for a bonus. Leaders cited a lack of complaints from schools as another indicator that educators were unaffected by the lack of publicized targets for

year 3. Nevertheless, one leader admitted that educators' indifference to the targets indicates a weakness in their overall approach to SPBP and the accountability system more broadly. For the targets to have mattered more to educators, this leader believed the DOE and UFT should have done a better job communicating to schools about the importance of targets and ways to reach them. He characterized this as an ongoing dilemma because "the accountability system was being formed and reformed" as SPBP was being implemented.

Also, in both school years, a sizeable minority of CC members reported that their school's Progress Report target had been too high. (We did not ask this question on the teacher survey.) In 2009, more than one-third (38 percent) of CC members thought their school's Progress Report target was too high. As one respondent noted in the open-ended comments in 2009, "We deserve the bonus, however the goals set for our school are not realistic." Similarly, in one of our case-study schools, staff viewed the Progress Report target as beyond their reach, which seemed to weaken the motivational effect of the bonus (as predicted by expectancy theory, as discussed in Chapter Two). At this school, interviewees frequently reported not paying attention to the bonus and were not disappointed in failing to win the bonus in year 1 because they perceived the likelihood of winning was very small (interestingly, this school ended up earning the full bonus in year 2).

One possible explanation for the 2009 results indicating perceptions that the targets were too high was that our survey data were collected prior to the announcement of the year 2 results, when a great majority of participating schools, particularly at the elementary and middle school levels, met their targets. Yet, even the results from the 2010 survey—administered the year after many more schools had won the bonus in year 2—indicated widespread feelings that targets were too high. As illustrated in Table 4.2, 43 percent of CC members reported that their school's Progress Report target had been too high.

Some case-study respondents expressed concerns about a lack of alignment among the multiple state and city accountability systems and metrics. For example, in one school, an administrator voiced frustration that their school was "in good standing" with the state, but that the city Progress Report system labeled them as "not doing well." The

opposite scenario sparked significant frustration on the part of several educators in another case-study school that was on the state's list of persistently low-achieving schools yet won the bonus two years in a row. Others in this school were baffled by a nearby school that was slated by the city to close but had received the bonus. One teacher explained:

> What I don't understand . . . Schools that the city deems failing, who have Fs and Ds, are getting bonuses. That just seems like "What's going on?" The city and state I understand because they use different formulas. But if this bonus program is in part from the city . . . what's wrong here? Why are you saying this school is an F and you want to close it and they just got their bonus?

And while this teacher understood that the school in question may have demonstrated progress and thus qualified for the bonus based on a different set of criteria, she maintained that the perception of contradictions was still problematic: "You're rewarding schools, but you're closing them."

**Concerns About Timing.** Staff also expressed concerns about the timing of the program**.** Eighty-four percent of teachers said it would be helpful if they could know at the start of the school year what the plans were for distributing the bonus. Similarly, in both survey years, approximately three-fourths of CC members said it would be helpful if they could know at the start of the school year what the plans were for distributing the bonus.[3] Although staff might have known that schools winning the bonus would receive the equivalent of $3,000 per staff member, they could not be sure this was the exact amount they individually would receive. Thus, many respondents in case-study schools thought it would be helpful to know at the outset the specific criteria the school planned to use to distribute the bonus pool (e.g., if it would be divided equally among staff, if specific individuals or job titles would receive more, or if staff could take specific actions during the year to

[3] The wording of this item differed slightly in 2009. On the 2009 CC survey, respondents were asked whether they agreed or disagreed with the statement "It would be helpful if staff knew at the start of the school year what the plans were for distributing the SPBP bonus award and how much additional pay an individual could earn." The latter part of this item was removed from the 2010 CC survey.

guarantee a greater share). Although NYCDOE instructed schools to formally convene their CCs and submit their distribution plans in the spring, about 20 percent of schools took the initiative within the first months of the school year to communicate to staff the criteria they planned to use for distributing funds.[4]

In one case-study school, the principal felt he and his staff were unable to be as strategic as they could have been had the CC been given time over the summer to develop and then communicate their distribution plan to staff at the start of the year. A high school science teacher from another school was also very critical of the timeline for year 1, questioning the fairness of a program in which staff members did not know at the start of the year what their targets were and the basis for awarding bonuses:

> That was one of the things I couldn't understand. How can you promise me a bonus and you haven't set the parameters until the end of the year? . . . Everybody didn't know if we qualified for the bonus or not almost until about half way through this year, then I think they knew the parameters . . . all the time the staff thought they had done better . . .we didn't know the targets at the beginning. . . . basically, when we knew that we didn't get the bonus, then I personally knew what the parameters were.

**Concerns About CCs.** Other staff raised questions about the composition of the CC**.** More than one-half of teachers (70 percent) and UFT-represented committee members (59 percent) questioned the

[4] The 20 percent (33 schools) is based on school-level measures calculated for 176 schools for which we had results. Our survey indicates that another 9 percent of schools had decided and communicated their distribution plans to staff in the middle of the year; 22 percent communicated the plans within the past month of taking our survey (in May and June); and 15 percent had decided but not yet communicated their plans to staff. The remaining 36 percent of schools had not yet decided on their plans at the time of our survey. In most cases, these school-level measures were based on the principal or other administrator's response. When no principal or administrator's response was available for a school, the response of the UFT member (if only one responded) or the common response of the two UFT members was taken as the school-level measure. School-level measures were not calculated for schools in which no principal or administrator responded and multiple UFT members' responses conflicted.

50–50 UFT–administrator representation on the CC and believed that UFT should have more than one-half of the representatives. In contrast, fewer than one-third of administrators believed UFT should have proportionately greater representation on the committee.

**Concerns About Staff Eligibility.** Finally, some respondents voiced concerns about staff eligibility for bonus awards. Administrators and UFT committee members disagreed on the question of whether program guidelines should be altered to ensure that only staff who "truly contribute to student success" receive a share of the bonus. Administrators were significantly more likely than UFT committee members to agree that guidelines should be altered in this way: 52 percent to 32 percent. Similar to UFT committee members, 38 percent of teachers overall believed the guidelines should be altered in this way. This is consistent with later items we report that indicate a strong preference among teachers and UFT committee members for an egalitarian distribution of the bonus share and greater willingness on the part of administrators serving on the committee to identify individuals they felt were deserving of a greater share of the bonus (see further details below).

Another concern expressed in several case-study schools was the exclusion of staff represented by employee union DC37—notably classroom aides (different from paraprofessionals represented by UFT), who many individuals believed contributed to school success and deserved a bonus share. Of course, UFT has no legal authority to negotiate for bonuses for these aides, who are part of a different union and covered by a different contract. Thus, even if UFT and NYCDOE had wanted to include DC37 staff, they would have had no authority to do so. In at least two case-study schools, staff decided that if they won the bonus they would pool funds from UFT-represented staff to provide at least some compensation to DC37 members (compensation that would not come directly from NYCDOE but in the form of checks written out personally from school staff).

**Differences in Opinions by School Bonus Status.** A few teacher opinions about SPBP differed depending on whether their schools did or did not win the bonus in year 2. A significantly greater percentage of teachers in the bonus schools reported that SPBP was fair to all schools (58 percent to 36 percent). Similarly, teachers from schools winning

the bonus were significantly less likely than teachers from nonbonus schools to agree that rewarding teachers based on test scores gains was problematic because it was hard to relate gains in student achievement to the work teachers had done (78 percent to 92 percent). Again, these differences can be interpreted in multiple ways. The more negative attitudes among teachers in the nonbonus schools could help explain why these schools did not earn the bonus (i.e., causation). The lower level of buy-in may have weakened the motivational effects of the bonus. However, another interpretation is that not winning the bonus soured teachers on the overall program and led teachers' attitudes to be more negative toward the program (i.e., consequence). It is also possible that the same teachers would have responded negatively in any situation, regardless of the program and the bonus outcome.

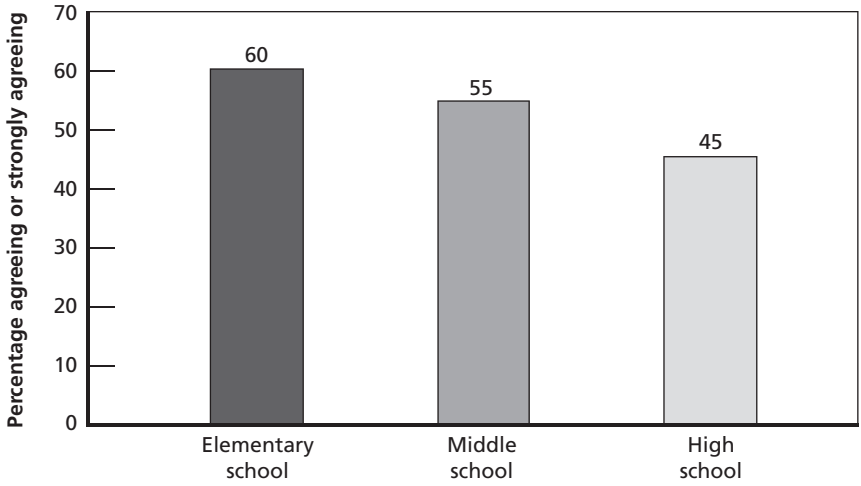
**Differences in Opinions by Level of Schooling.** Finally, some teacher opinions about SPBP differed by school level. In particular, elementary and middle school teachers were more likely than high school teachers to report that the SPBP method for awarding bonuses was fair to all participating schools (Figure 4.2). The differences between elementary and high school teachers were statistically significant.

### Although Many CC Members Recognized That a Higher Level of Performance Would Be Required to Win a Bonus in the 2009–2010 School Year, the Majority of CC Members Felt Fairly Certain Their Schools Would Do So

The 2010 CC survey asked respondents to rate their schools' chances of receiving the bonus based on the current year's student performance using a scale of 0 percent (no chance) to 100 percent (certainty). As illustrated in Table 4.3, on average, CC members reported that their schools had a 73- to 75-percent chance of receiving the bonus for their performance in 2009–2010.

This high level of certainty was simultaneous with an understanding many had that the bar for performance was being raised. As Table 4.4 illustrates, more than one-half of all CC members thought the criteria for receiving a bonus would require a higher level of performance, while one-third thought it would require a similar level of performance.

**Figure 4.2**
**Percentage of Teachers Agreeing or Strongly Agreeing That "The SPBP Method for Awarding Bonuses (Based on Progress Report Targets) Is Fair to All Participating Schools" (2010)**




RAND *MG1114-4.2*

**Table 4.3**
**CC Member Reports on Chances of Their Schools Receiving a Bonus Based on Current Year's Performance (2009–2010)**

| | Mean | Median |
|---|---|---|
| All CC members (n=496) | 74 | 80 |
| Administrators (n=195) | 73 | 80 |
| UFT staff (n=299) | 75 | 80 |

NOTES: Two respondents did not provide information about their titles.

These results provide some indication that, in SPBP schools, at least among CC members, one of the necessary conditions articulated by the theory underpinning pay-for-performance policy had been met: a high level of expectancy, a belief that individuals can improve student performance and meet program goals. If educators felt they had

**Table 4.4**
**Percentage of CC Members Reporting That the Criteria for Receiving a Bonus in 2009–2010 Would Require a Similar, Higher, or Lower Level of Performance Compared to 2008–2009**

| Level of Performance | All Committee Members | Administrators | UFT Staff |
|---|---|---|---|
| Similar | 35 | 30 | 40 |
| Higher | 64 | 69 | 59 |
| Lower | 1 | 1 | 1 |

NOTES: Statistically significant differences between administrators and UFT staff are indicated in boldface ($p < 0.05$). Total numbers and percentage calculations exclude those who responded "don't know."

no chance of winning the bonus, one would not expect to find that the bonus would have any motivational effect. Nevertheless, these results reflect only up to four individuals in each school and not the entire school faculty. Combined with the finding reported earlier that almost one-half of CC members believed their school's Progress Report target had been too high, it is still not clear how strong this level of expectancy really was in schools.

## Attitudes About the Bonus

This section describes staff members' attitudes about the financial bonuses, including reported desire to earn the bonus, level of satisfaction with the bonus amount, and thoughts about the salience of SPBP bonuses relative to other sources of motivation.

As noted earlier, in year 1, 96 schools won the full bonus, and 31 schools won partial bonuses; in year 2, 158 schools won the full bonus and eight schools won partial bonuses (see Table 2.2). The majority of survey respondents who said their schools received a bonus in year 2 reported personally receiving a share of the SPBP bonus.[5] Among UFT

[5] Among administrators, 101 respondents reported receiving the CSA bonus for the top 20 percent of administrators (and were thus not eligible for the SPBP bonus). Among the

committee members, more than one-half reported receiving between $3,000 and $3,199. Recall that, for administrators, amounts are predetermined by the agreement with the CSA (for a full bonus, principals received $7,000 and assistant principals $3,500; each received one-half that for a partial bonus). Nevertheless some administrators reported different amounts, perhaps because they too donated a portion of their awards to other, ineligible staff (as some case-study individuals did) or were reporting after-tax receipts.

### The Majority of Teachers and CC Members Reported a Strong Desire to Earn a Bonus Award

Sixty-four percent of teachers and more than three-fourths of CC members (78 percent) across schools reported a strong desire to earn a bonus award, with no significant difference in reports from UFT and administrative members on the CC. In case-study interviews, many UFT committee members felt that they had worked hard and therefore wanted and deserved the bonus. Others acknowledged that, although they did not buy into the assumption that the bonus would motivate them to work harder or smarter, any additional compensation was appreciated and thus desirable. As one teacher explained, "getting extra dollars in this economy is a plus."

### Respondents Whose Schools Earned Bonuses Reported a High Level of Satisfaction with the Bonus Amounts, but Some Dissatisfaction With the Ultimate Amount Received After Taxes—Dissatisfaction That Teachers Across all SPBP Schools Echoed

As Table 4.5 illustrates, the vast majority of teachers and CC members from schools that won the bonus in year 2 reported that they were satisfied with the amount they personally received. However, more than two-thirds of the teachers and UFT CC members and 60 percent of

---

67 other administrators who did not get the CSA bonus and were in schools that earned the SPBP bonus, 10 reported not receiving a bonus share. Among the 168 UFT CC members whose schools earned the SPBP bonus, 13 said they did not receive a bonus share. Case studies and interviews suggest that some of these individuals may inadvertently have been left off the distribution list or perhaps did not work at the school for the full year or full time and were thus excluded from the distribution plan.

**Table 4.5**
**Percentage of Teachers and CC Members from Schools That Earned Bonuses Agreeing or Strongly Agreeing with Statements About Bonus Awards**

| | | Compensation Committee | | |
|---|---|---|---|---|
| Statement | Teachers | All Members | Admin | UFT |
| I was satisfied with the amount I received | 79 | 81 | 80 | 81 |
| After taxes, the bonus amount I received felt insignificant | 73 | 66 | 60 | 69 |

NOTES: Statistically significant differences between Administrators and UFT staff are indicated in boldface ($p < 0.05$). Total numbers and percentage calculations for each item exclude those who responded "don't know." See Appendixes A and B for more detailed results on these items.

the administrators believed that, after taxes, the bonus amount they received felt "insignificant." Many individuals in case-study schools expressed a similar, seemingly contradictory opinion. On the one hand, they appreciated having some additional money ("anything is better than nothing," said one teacher), but once taxes were removed, they were disappointed at the actual amount they received. One teacher noted, "I was satisfied with the amount I was supposed to receive. I was not satisfied with the amount after taxes."

Some UFT-represented staff believed that administrators should not receive larger bonus amounts than teachers. One UFT member commented on the open-ended portion of the 2009 survey: "teacher bonuses should exceed what administrators receive. One reason is they facilitate the instruction that is necessary for students to perform well on the assessments." Even an assistant principal in one of the case-study schools complained that it was "insulting" to receive one-half of what the principal receives, noting that he performs far more than one-half as much work as the principal.

In general, however, most case-study respondents were happy to earn the money (one called it "found money") and reported purchasing Christmas gifts or other household items (e.g., a television), putting it toward their rent or mortgage, or paying off bills or debt. A few who

were philosophically opposed to SPBP donated their bonus share to charity.

Perhaps not surprisingly, UFT CC members who received a larger share of the bonus were significantly more likely to report higher satisfaction with the amount they received than those who received a smaller share. For example, 90 percent of those earning $3,000 to $3,199 and 86 percent of those earning $3,200 or more agreed or strongly agreed that they were satisfied with the amount they received; only 64 percent of those earning less than $3,000 did so. We did not find significant differences in reported levels of satisfaction among teachers based on their bonus shares.

Finally, among teachers across all SPBP schools—including those that won the bonus in year 2 and those that did not—44 percent reported that the bonus amount was not large enough to motivate extra effort. Consistent with these survey results, individuals in four case-study schools in year 1 reported that their bonus shares and the typical bonus share were not large enough, particularly after taxes, to lead them to make changes. For example, one teacher whose school did not win the bonus noted that "$3,000 is not a lot in the grand scheme of things. I spent $3,000 on classroom supplies, so it would go right back into the classroom anyway. I don't know that it would affect my performance." One teacher also felt strongly that the bonus was not enough money to risk the divisiveness that could result from a differentiated distribution plan. In three schools, some teachers suggested that other types of rewards in the form of supplies, materials, or even professional development would be more motivating than cash.

### Monetary Bonuses Rank High on a List of Possible Factors Motivating Teachers to Work Hard

When asked the extent to which a list of possible events would motivate them to work hard as a teacher, more than 60 percent reported that the possibility of "receiving a financial bonus for meeting school goals" motivated them to a moderate or great extent. Although a host of other possible factors ranked higher on this list—including observing student learning and improvement, achieving high marks on accountability measures, and external sanctions (see Table 4.6)—

financial bonuses nonetheless received significant acknowledgement as a potential motivating factor from a large share of teachers. Of course, even more teachers (three-fourths) noted that achieving a high Progress Report score was motivating, suggesting that the accountability system alone, without a financial incentive tied to the score, was serving a motivational purpose. These results are an important reminder of the intense, high-stakes accountability context in which SPBP was implemented. These other pressures—such as high scores on federal, state, and district accountability measures and avoiding sanctions—might have affected the ability of the SPBP bonus to motivate change. This may be particularly true if other policies in teachers' environments were better aligned to things that mattered more to teachers than the limited financial incentives of SPBP.

**Table 4.6**
**Percentage of Teachers Reporting Possible Factors Motivating Them a Moderate or Great Extent to Work Hard**

| Statement | Percentage |
|---|---|
| Seeing my students learn new skills and knowledge | 99 |
| Seeing my students improve their performance | 99 |
| Seeing myself improve as a teacher | 99 |
| Achieving or exceeding our school's Adequate Yearly Progress target | 77 |
| Receiving a high Quality Review score | 76 |
| Achieving a high Progress Report score/grade | 75 |
| Our school being designated for district or state intervention | 64 |
| Receiving a financial bonus for meeting school goals | 64 |
| Gaining recognition from my school colleagues for helping to meet our school goals | 64 |
| Gaining recognition from parents, the media, and/or district leaders for meeting our school goals | 62 |
| Our school being closed or phased out | 58 |
| Receiving public criticism due to not achieving our school goals | 44 |

NOTES: See Appendix B for more–detailed results on these items.

An analysis of three broad constructs derived from the survey data presented in Table 4.6—external reinforcement (e.g., receiving high Progress Report grades or Quality Review scores, receiving recognition from the public or colleagues for meeting school goals), external sanctions, and student learning—reveals that treatment and control-group teachers did not differ significantly in how they perceived motivating factors. As such, being a bonus school did not appear to reorient how teachers responded to possible motivators.

### Case-Study Interviewees Reported That Other Accountability Pressures and Intrinsic Motivation Were More Salient Than the Bonus

Across schools that did and did not win the bonus, principals, teachers, and other staff cared greatly about the city Progress Report, state report cards, quality reviews, the opinions of district leadership, and/or whether their schools would become subject to closure. Many were also highly attuned to parent and community opinions and concerned about their responses to any publicity about their schools and the schools' performance. Even though some of these accountability mechanisms were aligned with the bonus and SPBP is perhaps best viewed as a part of this larger accountability system, many individuals did not see that connection and in fact named these other pressures as stronger motivators than the financial bonus itself. This was particularly true in one-half of the case-study schools, where staff reported that concerns about being phased out, appearing on the state's list of persistently low-performing schools, and having negative outcome data publicized in the media were stronger incentives than the bonus for implementing school improvement efforts. For some, the threat of sanctions was more motivating. One UFT chapter leader believed the threat of "things happening to our school" was "very real" and much more effective than the potential to receive a bonus at driving her desire for the school to be successful. In another school, individuals who had experienced their comprehensive high school being closed and reopened as smaller schools felt strongly that the threat of another closure was more motivating than the bonus. A data specialist at yet another school echoed these sentiments:

> We do our job for our kids to give them the best education and make the most progress. Regardless of the bonus program, we are still held accountable to the city and the state. If our numbers go down, we're going to have people breathing down our necks, and I don't want to become a school in need of improvement or a SURR [Schools Under Registration Review, schools not meeting state standards and facing the possibility of closure] school. You know, that's scarier.

Others believed that the Progress Reports were motivating on their own, even without the bonus. They noted that regardless of the bonus, they were proud of earning high grades and worked hard to achieve and maintain them over time.

In many schools, staff members also typically attributed their hard work and their efforts to improve their practices to intrinsic motivations far above any external pressures or incentives. Many were quick to report that they did not enter teaching for the money, that money is fleeting, and that they chose their profession and sought to improve each year because of a commitment to helping children. One elementary school teacher from a school that won the bonus the previous year stated,

> Most teachers are doing the job they were paid to do, and they would do it regardless. I don't know if it was implemented to be an incentive to do better, but I don't see anyone here, especially myself, looking at it that way. The incentive is the job I do. To me I'm here to teach the kids, and I want to teach them. To me, it makes me feel good that they scored well because it makes me look good. I guess it's an ego thing.

### Conceptually, Most Individuals Interviewed in Case-Study Schools Viewed the Bonus as a Reward and Not as an Incentive

Consistent with the survey data, many case-study respondents across both years described the program and the value of the bonus as a means of recognizing staff for their hard work—work that they would be doing with or without the bonus. This was a major theme in four

schools in 2009, and individuals in all seven schools visited in 2010 expressed this sentiment. Across schools, staff frequently invoked a cake metaphor to describe the program, reporting that the bonus was the "icing" or "frosting" on the cake. A principal from a school that did not win the bonus the previous year reported,

> I don't think teachers were affected by SPBP. I think the teachers are going to work really hard. It's like icing on the cake. It's great if they get it, but they're not going to really change because of that. I didn't know about the program when I was planning school improvement efforts. . . . It doesn't affect how I or my staff approaches anything. Everyone is already giving 110 percent.

Similarly, a principal from a school that won the bonus explained, "It shouldn't be the end all be all. It should be the icing on the cake, because we did the job. This thing is to do the job. It shouldn't be the reason why we get the job done." "It's not the motivation," said another principal whose school won the bonus, "it's kind of like the reward."

Others described the bonus as "a nice pat on the back," but not the motivation for how they conducted their job. According to one teacher, the bonus sent a message that "we know it's hard, we see you've worked hard, you've really pushed it to the limits to get these kids the skills you need. Good job. We really appreciate it." Although we did find some exceptions (e.g., a 7th grade ELA teacher believed SPBP added some excitement and motivation to her work; a principal, assistant principal, and UFT chapter leader from one school who saw the bonus affecting teacher motivation and playing a key role in school improvement efforts), the majority of staff interviewed did not see the offer of a bonus as something that motivated them to work harder or in a different way. A teacher in a school that did not win the bonus said, "Every single year, bonus or not, I've been trying to improve on my strategies and methodologies." A teacher in school that won the bonus concurred, "I'm going to do what I have to do, regardless. That's what I get paid for, and I don't need brownie points." A colleague in this school commented, "it's not going to make me work harder."

## Communication and Understanding

This section examines the nature of communication about SPBP and the levels of program understanding that staff members in participating schools achieved. It describes NYCDOE and UFT communication strategies, CC member reports on information received and its adequacy, the methods of communication participating schools used, and some of the misunderstandings staff members conveyed to us.

### Although NYCDOE and UFT Leaders Acknowledged There Were Communication Problems in Year 1, There Appeared to Be Improvements in Subsequent Years

Both NYCDOE and UFT recognized that developing appropriate materials and presentations for year 1 was more difficult than they had anticipated. Promoting understanding of such details as how teachers would qualify for bonuses, how the CC was to operate, and what role the principal was to play proved to be a challenge. Commented one UFT official, "We were pretty shocked that, despite all the written instructions the schools had [at the beginning of year 1], they lacked very basic understanding, especially when it came to distributing the money." One of the program's principal funders concurred about the issue of communication:

> I think we could have been a lot more thoughtful about the importance of communication. . . .We could have invested a lot more time [and] money up front in making sure that . . . those people who were making the decisions [at the schools] were well informed and really aware of the full range of options available to them. Because I think the first year of the program, people didn't really understand what they were deciding.

NYCDOE and UFT quickly recognized that their year 1 joint orientations for eligible SPBP schools efforts were insufficient. As a UFT official noted, "We [NYCDOE and UFT] came to realize [in year 1] that the orientation was not adequate. Schools were not required to attend or send a representative, and there was generally poor understanding [of the program]." Confusion about the compensation pro-

gram was compounded by the relative newness of the Progress Reports at the heart of determining bonus eligibility. Both NYCDOE and UFT acknowledged that Progress Reports can be difficult to understand. Said a NYCDOE official, "I think the [compensation] program would be a little bit more on solid ground if the Progress Reports had been around for, say, five years and had time to work out a lot of their issues." Seconded a UFT official, "Having people [at the schools] understand how to make the [Progress Report] targets was a failure [in year 1] of both the UFT and the [NYCDOE]."

Lessons learned from SPBP's year 1 orientations were used to improve year 2 orientations. In year 2, NYCDOE and UFT required eligible schools that might be interested in participating in the program to attend the orientation sessions. Both the school system and the union reported that, as a result of this change and improvements in the presentation of material about the program, orientations went more smoothly in year 2. They believed that, in year 2, schools had a better understanding of SPBP requirements.

In year 3, program leaders believed that, because no new schools were participating, they did not need to hold additional meetings at the start of the school year. Leaders believed most schools were fairly well oriented to the program and the Progress Report system.

### While Teachers Tended to Receive Information About the Program from UFT, Most CC Members Received Information from a Variety of Sources

The NYCDOE-UFT communication plan for SPBP called for sharing joint written materials about the program with school staff. Both organizations adhered to this plan, developing materials together and cosponsoring all formal meetings for CCs. Within schools, UFT representatives were one of the main conduits of information. As illustrated in Table 4.7, the majority of teachers reported receiving information about SPBP from UFT representatives or chapter leaders (72 percent) and their school administrators (55 percent). These were also the main sources of information for UFT staff on the CC, but over one-half of the CC members who responded in 2010 received information from NYCDOE, UFT central office and representatives, and school-level

**Table 4.7**
**Percentage of Teachers and CC Members Receiving Information About the Schoolwide Performance Bonus Program**

| | | Compensation Committee | | | |
|---|---|---|---|---|---|
| | Teachers | Administrators | | UFT Staff | |
| Source | 2009–2010 | 2008–2009 | 2009–2010 | 2008–2009 | 2009–2010 |
| UFT central office[a] | 30 | **61** | **48** | **74** | **58** |
| My school's UFT rep./chapter leader[a] | 72 | **78** | **65** | **90** | **77** |
| NYC Department of Education | 30 | **92** | **87** | **68** | **59** |
| School administrators | 55 | **73** | **65** | **82** | **69** |
| NYC Teaching Fellows or Teach for America | 3 | NA | 2 | NA | 2 |

NOTES: Statistically significant differences between Administrators and UFT staff are indicated in boldface ($p < 0.05$). Total numbers and percentage calculations for each item exclude those who responded "don't know."

[a] Statistically significant differences across years were observed on these CC items. See Appendixes A and B for more detailed results on these items.

leaders. Consistent with leader reports of reduced emphasis on informing staff about SPBP in year 3, fewer CC members reported receiving information from each source.

### CC Members Generally Reported Receiving Sufficient Guidance in a Number of Areas

For example, in each of the six areas related to understanding and implementing various aspects of the program (e.g., how to interpret Progress Reports, the options for distributing bonus awards to staff), more than 80 percent of respondents reported receiving guidance (see Table 4.8). Among those who did so, one-half or more believed the information was sufficient and met their needs, and very few wanted more guidance in any area. Although the results presented in Table 4.8 represent the 2009–2010 school year, they differ only slightly from the results from 2008–2009.

Among committee members, UFT staff were less likely than administrators to receive guidance in all but one of the six areas listed in Table 4.8, such as guidance about how to identify strategies that would maximize their chances of meeting targets and how to interpret Progress Reports. Nevertheless, UFT staff only had slightly less positive perceptions than administrators about the adequacy of the information they received in one topic area: how to interpret Progress Reports.

These differences may be due in part to the method UFT and NYCDOE used to communicate with school staff. For example, throughout the school year, the NYCDOE administrator overseeing SPBP communicated directly with principals but was unable to directly contact teachers serving on the committees. Because most CC members were not elected until the end of the year, there was no central list of UFT-represented CC members and email addresses to be used for communicating early in the school year. Also, some individuals we interviewed thought that UFT and NYCDOE had agreed that UFT would communicate directly with UFT-represented staff on the CC and that NYCDOE would communicate with administrators on the CC.

**Table 4.8**
**Percentage of CC Members Reporting on Receipt and Sufficiency of SPBP Guidance (2009–2010)**

| Area of Guidance | Respondents | Did not receive | Received, and It Was[a] | | |
|---|---|---|---|---|---|
| | | | Sufficient: Met my needs | Somewhat sufficient: Partially met my needs | Not sufficient: I want more guidance |
| How to identify strategies that maximize our chances of achieving a high Progress Report grade | All | 20 | 72 | 24 | 4 |
| | Admin | **15** | 74 | 22 | 4 |
| | UFT | **23** | 70 | 25 | 4 |
| How to communicate the goals of the SPBP to all staff | All | 16 | 75 | 21 | 5 |
| | Admin | **11** | 80 | 15 | 5 |
| | UFT | **19** | 71 | 24 | 5 |
| The criteria for winning a school bonus | All | 15 | 72 | 22 | 6 |
| | Admin | **9** | 71 | 21 | 8 |
| | UFT | **18** | 72 | 23 | 5 |
| How to interpret Progress Reports | All | 15 | 73 | 24 | 4 |
| | Admin | **9** | **80** | **18** | **2** |
| | UFT | **20** | **67** | **27** | **5** |
| How to achieve consensus among our Compensation Committee concerning our school's award distribution plan | All | 15 | 82 | 14 | 4 |
| | Admin | 11 | 85 | 11 | 4 |
| | UFT | 17 | 81 | 15 | 4 |

**Table 4.8—Continued**

| Area of Guidance | Respondents | Did not receive | Received, and It Was[a] | | |
|---|---|---|---|---|---|
| | | | Sufficient: Met my needs | Somewhat sufficient: Partially met my needs | Not sufficient: I want more guidance |
| The options for distributing SPBP bonus awards to individual staff members | All | 12 | 82 | 12 | 6 |
| | Admin | **7** | 82 | 11 | 7 |
| | UFT | **15** | 82 | 13 | 4 |

NOTES: Statistically significant differences between administrators and UFT staff are indicated in boldface ($p < 0.05$). Total numbers and percentage calculations for each item exclude those who responded "don't know."

[a] Percentages in these three columns were calculated based on those who reported they had received listed guidance about SPBP. See Appendixes A and B for more detailed results on these items.

### Although Levels of Understanding About SPBP Were Reported to Have Improved Over Time, Communication Problems and Misunderstandings About the Program Remained

On a positive note, in year 2, more than three-fourths of all CC members (78 percent of administrators and 80 percent of UFT staff) reported that they understood the Progress Reports better in 2008–2009 than they had in 2007–2008. In year 3, the frequency of reported improvement in understanding declined slightly: two-thirds of all CC members (64 percent of administrators and 68 percent of UFT staff) reported that they understood the Progress Reports better in 2009–2010 than they had in 2008–2009. These differences over time are not surprising, given that, by year 3, educators had several years of exposure to the Progress Reports. In year 1, SPBP was announced later in the year, and the Progress Reports were still quite new. One would then expect a bigger jump in understanding from years 1 to 2 than years 2 to 3.

Consistent with this pattern, the evidence suggests that teachers' understanding of Progress Reports may also have improved over time. In 2009, almost one-half of CC members said that there was a lack of understanding among teachers at their schools about the school's Progress Report and the factors that had contributed to the school's score—although UFT members were significantly more likely to have this opinion than administrators (51 percent to 32 percent), as were respondents from nonbonus schools compared to bonus schools (54 percent to 38 percent). In 2010, 40 percent of CC members reported that not all teachers at their school had understood the school's Progress Report and the factors that had contributed to the school's score. Although the frequency declined, the results nonetheless indicate that a significant minority of CC members believed their teacher colleagues did not adequately understand these accountability metrics.

In fact, 84 percent of teachers in 2010 reported having a strong understanding of their school's Progress Report and contributing factors. Nevertheless, a sizeable minority of teachers reported not understanding several aspects of the program very well, such as the amount of funding a school would receive if it met 100 percent of the target, the criteria for receiving a partial bonus, and the source of funding for the program (see Table 4.9).

**Table 4.9**
**Percentage of Teachers Reporting Level of Understanding of SPBP Elements (2009–2010)**

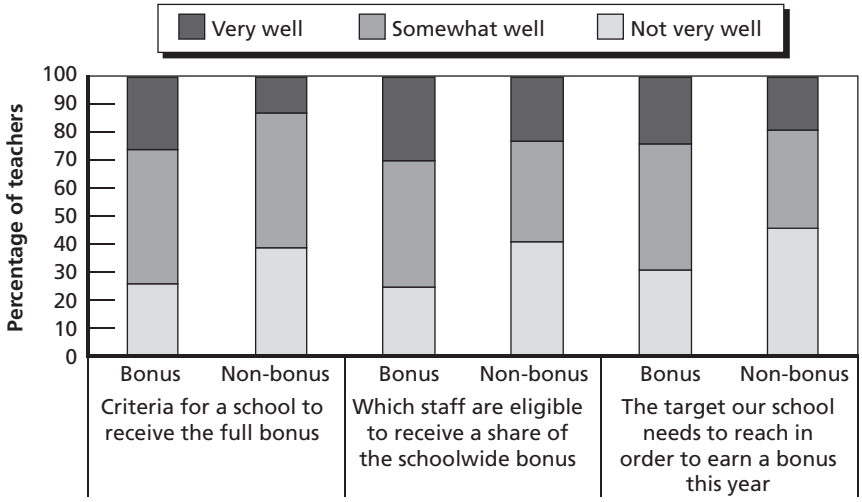
| Variable | Not Very Well | Somewhat Well | Very Well |
|---|---|---|---|
| Which staff are eligible to receive a share of the schoowide bonus | 27 | 45 | 28 |
| Criteria for a school to receive a full bonus | 29 | 48 | 23 |
| How a school compensation committee decides on a distribution plan for the bonus | 34 | 40 | 26 |
| The target our school needs to reach in order to earn a bonus this year | 34 | 43 | 23 |
| The amount of money our school would receive if we met 100% of our target | 36 | 38 | 26 |
| Criteria for a school to receive a partial bonus | 41 | 44 | 15 |
| The source of funding for this program | 56 | 34 | 10 |

NOTES: See Appendix B for total number of respondents on each question.

Teachers from schools earning the bonus in year 2 were significantly more likely than those in schools that did not earn the bonus to report understanding three aspects of the program: criteria for schools to earn the bonus, which staff were eligible to receive a share of the bonus, and the target their school needed to reach to earn the bonus (Figure 4.3). Similarly, a significantly greater percentage of bonus group teachers reported that they had a strong understanding of the Progress Report and contributing factors than their nonbonus group colleagues: 86 percent versus 75 percent.

As noted earlier, these differences could be interpreted in multiple ways. The lower level of understanding among teachers in the nonbonus schools could help explain why these schools did not earn the bonus (i.e., causation). In this sense, the findings help support the idea that motivational effects of the bonus were weaker when individuals did not fully understand the program and criteria used to allocate bonuses. Alternatively, not winning the bonus could have increased the

**Figure 4.3**
**Percentage of Teachers Reporting Level of Understanding of SPBP Elements (2009–2010)**




RAND *MG1114-4.3*

negative responses to questions about the program, or the same teachers might have responded negatively in any situation, regardless of the program and the bonus outcome.

The case studies revealed even greater communication problems and misunderstandings. Interviewees indicated problems with how they received information about the program and, in turn, their understanding of multiple facets of the program. While some admitted to not understanding the program, others indicated this through the misinformation they conveyed when responding to our questions. Typically, principals had the most accurate understanding of the program, followed by UFT chapter leaders and CC members, followed by teachers and then nonteachers.

Staff generally relied on the principal and UFT leader for information about the program, but these individuals often did not disseminate the information effectively or helped perpetuate misinformation. Although these communication problems were particularly acute in year 1—a problem acknowledged by leaders in many of our interviews

and an area identified as greatly improved by many case-study respondents—problems lingering in years 2 and 3. When asked to advise NYCDOE and UFT on ways to improve the program in the future, at least one individual (and in some cases, many more) in all but one of the 14 schools visited in 2009 and 2010 cited a need to improve communication or requested more information. For example, some individuals wanted more written materials explaining the program, while others requested that well-informed leaders from UFT headquarters visit schools to directly inform staff about the program, rather than filter communication through UFT chapter leaders. Some wanted more information about the targets and bonus criteria, while others wanted more specifics on the CC process.

One consequence of the reported communication deficiencies was that staff developed many misperceptions about SPBP. In almost every school visited, we met with staff members who held misperceptions regarding one or more of the following:

- whether their school was participating in the program or the number of years the school had participated
- whether they had won the bonus and/or whether it was the full or partial amount
- the criteria for awarding bonuses (e.g., some were unaware that school environment survey results and attendance were included in the calculations) and allowable options for awarding bonus amounts (e.g., one school was unaware that they could assign $0 to individuals)
- the bonus amount (e.g., one teacher thought each school received $5,000 total)
- which staff were eligible (e.g., some thought only classroom teachers could receive shares)
- rules guiding development and dissemination of the distribution plan (e.g., one interviewee thought the entire school staff had to vote on the plan; another believed the CC could not share the plan with staff)
- the source of funding for the program (e.g., one teacher reported that SPBP was privately funded and that mattered to her greatly)

- which organizations ran the program (e.g., some were unaware that UFT was a cosponsor)
- whether schools were required to vote annually to participate.

These misunderstandings were especially widespread in three of the seven schools visited in 2009 and in one of the seven schools visited in 2010. For example, in one of these schools, staff believed that schools needed to earn an A grade two years in a row to receive the bonus and were not aware that the main criteria were based on the Progress Report targets. This misperception had spread throughout the school and was repeated to us in most interviews. The principal in this school admitted, "I would like more clarity about what you need to do to earn the bonus. . . . Another school I know about got an A 2 years in a row, but they weren't 'Well Developed' so they didn't get the bonus." In another school, three of the seven teachers interviewed were unaware that they were participating in the program, and others did not know which individuals in the school were eligible to receive bonuses or how much was at stake (one teacher admitted that she didn't even know the "ballpark"—"is it thousands, hundreds, $20? . . . I really have no idea."). Perhaps some of the misinformation in these particular case-study schools could be traced to school leaders. In one school, the principal had taken on leadership midyear, which could help explain some of the communication problems. In another school, the principal was knowledgeable about SPBP, but the UFT chapter leader lacked a strong understanding about the program components.

These misperceptions may be helpful in understanding some of the attitudes reported earlier in this monograph, particularly the degree to which staff viewed the bonus as a meaningful motivator. They also raise important questions about the underlying theory of action and implementation of the program: (1) If staff members did not understand the criteria, how would they know where to direct their efforts to achieve the bonus? (2) If staff members did not know the amount at stake, how could they gauge whether the payoff was worth the effort? and (3) If staff members did not know that they were eligible or that their school was participating, how could the program influence their attitudes or behavior? We will return to these topics later.

## Summary and Discussion

NYCDOE and UFT leaders, teachers, school CC members, and case-study school staff conveyed mixed attitudes and levels of understanding about the program. With regard to overall attitudes about the program, we found that

- While most respondents appeared to be aware of the program, SPBP was not central to staff conversations.
- Overall support for the program was quite strong. Most educators truly appreciated the recognition and financial reward for their efforts that the bonus program represented.
- Educators' attitudes toward the program, however, indicated both support and some areas of concern. Although most CC members and just over one-half of teachers thought the program was generally fair to participating schools, many raised questions about the measures used to determine bonuses (the majority of teachers and CC members felt criteria relied too heavily on test scores), the timing of the program (three-fourths or more of teachers and CC members said it would have been helpful to know plans for distribution at the start of the year), and the makeup of the committees (more than one-half of teachers and UFT committee members wanted more than 50-percent representation). Some also felt the Progress Report targets were too high.
- Although many CC members acknowledged that a higher level of performance would be needed to win the bonus in 2009–2010 than in 2008–2009, a large majority felt certain their schools would receive the bonus.

As for attitudes about the bonus itself, we found that

- The majority of teachers (64 percent) and CC members (78 percent) expressed a strong desire to earn the bonus.
- Although the majority of teachers and CC members who received a bonus were satisfied with the amount, more than one-half said that, after taxes, it felt insignificant. Across all SPBP schools,

44 percent of teachers reported on surveys that the bonus was not large enough to motivate extra effort.

- More than one-half of teachers (64 percent) reported that the possibility of receiving a bonus motivated them to a moderate or great extent to work hard as a teacher. Three-fourths, however, reported that achieving a high Progress Report score alone (without a bonus attached) motivated them to a moderate or great extent to work hard as a teacher.
- Conceptually, case-study respondents viewed the bonus as a reward, not an incentive. They reported that the bonus made them feel appreciated for their hard work but that this work would have been undertaken with or without the bonus.

Finally, the reports on communication and understanding were also somewhat mixed:

- Although most stakeholders reported improvements in communication over time, many misunderstandings and misperceptions appeared to linger.
- Although 40 percent of CC members reported that teachers did not understand Progress Reports and the factors contributing to these ratings, most teachers themselves reported having a strong understanding (84 percent). But a sizeable minority of teachers (more than one-third) reported not understanding several aspects of the program, including the criteria for receiving partial bonuses, the amount of money their school would receive if they met their target, the source of funding for the program, the target their school needed to reach, and how committees decided on distribution plans. In case-study schools, many individuals conveyed misperceptions about various aspects of the programs (e.g., whether their school was participating or the number of years they had participated, if they had won the bonus, how bonuses are awarded, allowable options for distribution, the bonus amount), and many identified communication as an area in need of improvement.

- On a positive note, one-half or more of CC members believed the guidance they had received about the program was sufficient and met their needs, and two-thirds said they had a better understanding of Progress Reports in year 3 than in the previous year.

In the aggregate, these findings indicate both promise and potential problems with program implementation and the prospect of SPBP having intended motivational effects on participating school staff. On the one hand, there is evidence that many supporting conditions were present in these schools, including a high level of buy-in and support for the program overall, a perception that the program was fair to all schools, a strong desire among staff to earn the bonus, and a belief that financial bonuses can be motivating. On the other hand, not all of the conditions appeared to be supportive of SPBP and its intended effects. Evidence of misunderstandings about key program components, concerns about the size of the bonus, and the ranking of other motivational factors higher than a financial bonus may have lowered the motivational potential of the bonus and limited SPBP's effects.

CHAPTER FIVE

# Implementation of the Schoolwide Performance Bonus Program: Compensation Committee Process and Distribution Plans

This chapter explores the implementation of one of the defining features of the New York City SPBP: the CC. As noted in Chapter Two, NYCDOE and UFT believed the CC would ensure that decisions about allocating bonus awards among staff would be school-based, allowing each school committee to decide whether they wanted to differentiate based on job title or levels of performance or whether they instead valued a more egalitarian distribution plan. Some leaders also believed that the possibility of schools differentiating bonus distributions based on individual performance might enhance the motivational effects of the bonus and lead to even greater positive outcomes for students. To these ends, SPBP outlined several guidelines regarding the operation of CCs, including the composition (four members that would include the principal, principal-designee, and two members selected by UFT-represented staff) and requirements for decisionmaking (no votes taken, consensus only) and criteria for differentiating bonus awards (seniority is barred). Some of these guidelines—such as those regarding membership and consensus-based decisions—were intended to ensure that neither administrators nor UFT-represented staff controlled the results of the CC process.

Our goal in this chapter is to examine this critical design feature and component of the theory of action underlying SPBP. The objective is an understanding of how the CC arrived at the planned award distributions and what those distributions were. This process

is important for understanding how SPBP played out and whether or not the plans were consistent with what the program developers had envisioned. It is also useful for understanding how school-based teams in other pay-for-performance programs that use a similar structure for determining awards might make decisions about compensation. This chapter also provides general lessons about the process of high-stakes decisionmaking by school-based teams.

As we will show, most schools followed SPBP guidelines regarding the CC process, and most CC members reported that the decisionmaking process was fair, collegial, and inclusive and that achieving consensus was easy. Nevertheless, some teachers and case-study school respondents reported some potential problems with the process. In the end, most CCs developed award distribution plans in which the vast majority of staff in each school received the same award amount and in which the primary factor for differentiation of awards was whether a staff member had worked at a school part time or part year. Consequently, if the ability of staff members to earn larger or smaller amounts based on personal performance was a key incentive for improving schools and student outcomes, as some leaders had hypothesized when designing SPBP, for the most part, the CC system did not support this element of the theory of action.

The research questions answered in this chapter are

1. How did the CC function?
    b. Who was on the committee, and how was this decided?
    c. How did the committee arrive at decisions?
    d. What did the CC member experience during the process?
    e. How did other staff members in the schools view the CC process?
2. What were the results of this process?
    a. What were staff preferences for distributing the bonus among staff?
    b. What factors did the CC report using to determine how to distribute bonuses?
    c. Did these include personal behaviors and performance that might be associated with motivating individual behaviors?

d. What were the resulting distribution of awards?
e. Were awards differentiated?
f. How did staff members respond to bonus distributions and distribution plans?
   - In schools winning the bonus, how did staff respond to bonus distributions, particularly differentiated distributions?
   - How was the planned distribution communicated to the staff?

## Compensation Committee Process

This section addresses the first broad research question, describing who served on the CCs and how they were selected as members. In the next section, we will examine the nature of decisionmaking, including the level of ease or difficulty the committees had achieving consensus.

### Although the Compensation Committees Were to Include Two UFT Selected Members, the Principal, and the Principal's Appointee, the Actual Number of UFT Members Varied Somewhat

Table 5.1 summarizes the membership statistics for the 2010 CCs by primary job classification and method of becoming a member. About 60 percent of all CC members were represented by UFT, and 41 percent were CSA members (the values in 2009 were 56 percent UFT and 44 percent CSA). The breakdown was not 50-50 because of the different choices principals made in appointing their one representative to the committee. In most schools, the principal appointed another administrator (e.g., the assistant principal), but in some schools, the principal appointed a UFT-represented staff member to the committee.[1] More than one-half of other UFT committee members were elected or appointed to serve by their UFT-represented colleagues. Approximately 20 percent of teachers and other UFT members were

[1] In some cases, the assistant principal of a school is officially classified as a teacher and is represented by UFT.

**Table 5.1**
**CC Members, Percentages by Role and Method of Becoming a Member (2009–2010)**

| Statement | Principal (N=122) | Other Admin (N=104) | Teachers (N=253) | Other UFT Members (N=78) | Total (N=557) |
|---|---|---|---|---|---|
| I am the principal and am required to be on the committee | 100 | N/A | N/A | N/A | 22 |
| I was appointed by the principal to serve on the committee | N/A | 91 | 8 | 15 | 23 |
| I was elected by UFT members to serve on the committee | N/A | 2 | 54 | 41 | 31 |
| I was appointed or designated by UFT members to serve on the committee | N/A | 2 | 9 | 13 | 6 |
| I am the UFT chapter leader and it was assumed that I would serve | N/A | N/A | 21 | 20 | 12 |
| I volunteered | 0 | 3 | 8 | 12 | 6 |
| Others | 0 | 2 | 0 | 0 | 0 |
| Total | 100 | 100 | 100 | 100 | 100 |

chapter leaders who were not elected and who assumed that, given their leadership position, they were expected to serve on the committee. A small group of CC members reported that they had volunteered to serve on the committee. These patterns are very similar to those we observed in 2009.

Regardless of their role on the committee, survey data indicate that CC members were experienced on several levels. First, three-quarters of all members in 2009–2010 (86 percent of administrators, 68 percent of UFT staff) had served on their school's CC during the prior school year (2008–2009). These percentages were up slightly from 2008–2009, in which 79 percent of administrators and 66 percent of UFT staff CC members had also served in the prior year. Committee members also tended to be veterans both in their current positions and in their schools. The majority of respondents had been working in their

current schools for more than a year (only 2 percent of CC members reported that this was their first year at their schools), and most had been in their positions for at least several years. Again, this was similar to 2008–2009, in which 4 percent of CC members were new to their schools.

### The Process for Determining Who Served on the CC Varied Considerably Among Schools

The case-study schools demonstrated differences in the processes for selecting CC members. A few of these schools adopted very formal and purposefully democratic systems for electing members—paying careful attention to formal voting procedures and to identifying representatives whom they trusted and believed would be fair. In one case-study school that differentiated the bonus distribution in past years, almost every interviewee indicated putting careful thought into whom they elected to serve on the committee, noting that they selected individuals because "they represented my ideas" or "I trust her judgment." Other case-study schools made sure to elect or appoint members that promised to represent what they felt was the majority view supporting equal distribution of the bonus.

In contrast, the procedures at many other schools were less formal. Interviewees in these schools reported that UFT-represented members of the CC were typically not selected because the staff in the school had an explicit understanding of the elected individual's position regarding the allocation of bonus awards. In general, only a few individuals had volunteered to serve on the committee, and they had not publicly stated their views prior to the vote (although this did occur in some schools, particularly those with strong preferences for an egalitarian distribution plan). In fact, in one case-study school, the principal appointed all CC members—a clear violation of the intent of SPBP.

The process was also highly charged in some schools. For example, in one case-study school, the UFT-represented staff held a revote after one CC member who had been elected created some controversy by raising the idea of differentiating the payout. This individual later resigned from the CC in hopes of easing tensions in the school.

Case-study schools also varied in choices involving UFT chapter leaders and principal-designees. In some schools, the UFT chapter leader was a CC member—in some cases, they were formally elected; in others, it was just assumed they should serve. In still other cases, the chapter leader purposefully chose not to serve on the committee. In two cases, the UFT leader felt serving on the committee could lead to potential conflicts of interest in representing staff interests. In another case, the UFT leader wanted to give other staff an opportunity to lead and be exposed to school decisionmaking. A UFT leader in one school later regretted serving on the committee and felt "caught in the middle" of having to defend a decision to differentiate the bonus payout to staff she was supposed to represent. "I'm going to have a whole bunch of angry teachers around me," She explained, "You are supposed to try to win the trust of the staff. [With] this, I'm losing the trust." Although most case-study principals designated an assistant principal to serve on the committee, a UFT staff member (a data specialist) was appointed in at least one instance. In one large school, the principal rotated a different assistant principal onto the committee each year.

Interestingly, in a few schools, staff members reported learning over time how important it was to elect the "right people" to represent them on the committee, admitting that they had not thought seriously about the selection process in year 1. For example, in one school, UFT-represented staff reported learning the importance of electing individuals to serve on the CC who could "stand up" to administration and make their voices heard.

If the intent of the CC was to ensure that decisions regarding bonus distribution among staff were school based and not driven by any particular stakeholder group, then CC membership was one important element supporting this objective. Overall, CC reports indicate efforts very much aligned with SPBP guidelines and intent, and provide few indications that the selection process created major power imbalances. The data in some case-study schools, however, suggest that who served on the CC and how they arrived on the CC may have affected decisionmaking dynamics (e.g., one person dominating the process or CC decisions being perceived as not representing the will of

the majority). As discussed below, however, survey data indicate these problems were not widespread.

### Most CC Reported the Decisionmaking Process Went Smoothly with Few Problems; a Minority of Schools Encountered Some Difficulty

As shown in Table 5.2, CC members felt that the decisionmaking process was fair, collegial, and inclusive. The overwhelming majority reported that their opinions and ideas were valued by fellow members, that members had an equal say in determining the distribution plan, that they trusted their colleagues on the CC to make decisions in the best interest of all staff, and that it was easy for the committee to achieve consensus (as required by the policy). Further, very few (14 percent) reported that some committee members dominated discussions.[2]

We found a similar pattern in the majority of case-study schools where CC members reported that achieving consensus was easy and that little conflict emerged. In some cases, respondents indicated that the requirement to reach consensus appeared to foster compromise. In these schools, individuals described agreeing to final decisions even if they were not entirely in favor of the decision in "the spirit of working together" and in recognition that failing to do so "means no one gets anything." For example, in one middle school with a differentiated distribution plan, administrators and UFT-represented staff on the committee cited several instances of conceding on certain points about which they felt strongly to reach consensus and not lose out on the opportunity to earn the bonus funds. "No one got exactly what they wanted," explained one UFT CC member, "I was happy, because they [administration] gave in too."

In other schools, staff reported little difficulty achieving consensus because they purposefully did not open up discussion about differentiating the distribution. These individuals believed it was not worth the risk of creating "disharmony" among staff.

---

[2] Analysis of responses in a given school indicates fairly unanimous views. In schools with more than one respondent, 80 to 90 percent of respondents reported unanimous answers to the questions listed in Table 5.2.

**Table 5.2**
**Percentage of CC Members Agreeing or Strongly Agreeing with Statements About the Committee Process (2009–2010)**

| Statement | All Members | Admin | UFT |
|---|---|---|---|
| I feel that my opinions and ideas were valued by other members of the committee | 95 | 95 | 95 |
| All members had an equal say in determining the final distribution plan | 94 | 92 | 95 |
| I trust committee members to make decisions in the best interest of all staff | 94 | 95 | 93 |
| It was easy for the committee to achieve consensus on how to distribute the bonus | 89 | 89 | 89 |
| [For those who were on the committee last year:] The process of achieving consensus on the development plan was easier this year compared to last year | 60 | 55 | 64 |
| Some committee members tended to dominate the discussion during the meeting(s) | 14 | 13 | 15 |

NOTES: Statistically significant differences between Administrators and UFT staff are indicated in boldface ($p < 0.05$). Total numbers and percentage calculations exclude those who responded "don't know/NA" to each item. See Appendix A for more detailed results on these items.

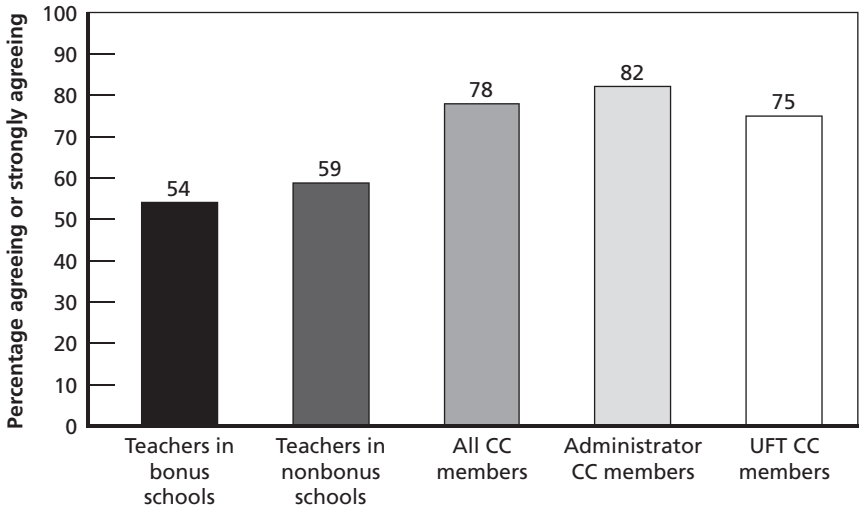
In several schools visited in 2010, we were told that the process of achieving consensus became easier over time. By year 3, these schools had already developed ground rules and understood the program and requirements. As shown in Table 5.2, 60 percent of CC members agreed that the process of achieving consensus was easier in year 2 than in year 1. (It is not clear whether the other 40 percent, who disagreed, believed the process was harder than or just the same as the prior year.) Interestingly, the increasing ease of reaching consensus builds on what was already described as an easy process the prior year. In 2009, 51 percent of CC members reported that it was "very easy" to achieve consensus on how to distribute the bonus among staff in year 1, and 42 percent described it as "somewhat easy: it required some discussion." Only 7 percent said it was "not easy: it required a lot of negotiation."

Despite these positive reports, there is evidence that some schools struggled with the CC process. Slightly less than one-quarter (22 per-

cent) of CC members disagreed or strongly disagreed that teachers' preferences were taken into account when developing the distribution plan—suggesting that not everyone felt the decisionmaking process was inclusive. Teachers were even more likely to question whether teacher preferences were factored into distribution plan decisions. In fact, as Figure 5.1 illustrates, only about one-half of teachers in both schools that won the bonus and schools that did not win the bonus felt that teachers' preferences were taken into account when developing the plan, indicating that almost one-half felt otherwise.

Case-study data illustrate several ways in which a minority of schools struggled with the decisionmaking process. In a few case-study schools, the assumption that equal numbers of administrator- and UFT-selected members and the requirement to achieve consensus would result in an "even playing field" was not necessarily realized. In some of these cases, principals were reported to have the final say on all decisions. There also were reports that the power differential between

**Figure 5.1**
**Percentage of Teachers and CC Members Agreeing or Strongly Agreeing with the Statement "Teachers' Preferences Were Taken Into Account When Developing the Distribution Plan" (2010)**




RAND *MG1114-5.1*

the administration and staff played a role in the deliberation and led to some UFT staff feeling that they were pushed into decisions with which they were not entirely comfortable. As one teacher committee member explained,

> We thought we were kind of strong-armed. It was like, the administrator came in with a game plan, and we didn't. So it's like the kids say, "too bad for us." We should have had our plan, something in mind that we wanted to do and stood up for it, but we didn't. . . . I have to be honest, a few people stopped talking to me. I felt bad because I'm here a long time. But you know what, they are talking to me again. So hey, it was a learning situation. I guess I wasn't as strong as I thought I was. Because when I took the position [on the Compensation Committee], and they voted me in, I thought I would be able to stand up, but I wasn't, I was kind of like "She's an administrator, and we'll kind of go with how she felt." And we shouldn't have.

In one case-study school, the power differential played out in more subtle ways. UFT CC members reported feeling uncomfortable with the principal's idea of giving teachers receiving unsatisfactory (U) ratings no share of the bonus pool but agreed to do so because they believed it would only affect a few individuals. Later, when submitting the final distribution plan to NYCDOE, they discovered that a substantial number of individuals had received U ratings and thus no bonus share. These UFT CC members were disturbed by the outcome and admitted they would have "fought harder" against the decision to award $0 had they known the number of U-rated teachers. This asymmetry of information clearly privileged the principal in the decisionmaking process. The inclusion of evaluation data as criteria for differentiating the bonus raises a host of other concerns as well, including questions about confidentiality. We were told on several occasions that CC members felt uncomfortable discussing teacher evaluations in committee.

Nowhere is the struggle around decisionmaking more clear than in one school (not in our case-study sample) that had to drop out of the program at the end of year 1 because the CC could not reach consensus

on its bonus distribution plan. Later, it became known that the school met 100 percent of its performance target and would have received more than $320,000 if the committee had agreed on the distribution scheme.

## Distribution Plans: Staff Preferences and Final Plans

When determining awards, the key decision of each committee was whether or not to differentiate awards. The debates on equal versus differentiated awards involved both the range of awards and decisions about which factors might be used to differentiate the awards. This section examines the preferences CC staff and teachers reported they had about differentiated as opposed to nondifferentiated bonus distribution, then goes on to an analysis of the plans submitted to NYCDOE and the factors reported to have been used in determining these plans.

As noted earlier, in the spring of each school year, each CC submitted a key piece of information to NYCDOE online: the amount of money, should the school win the bonus in the fall, that would go to each full-time UFT-represented staff member and any other staff member in the school that the committee determined should receive an award. Although the total amount awarded to a school was based on an equal dollar amount per full-time UFT-represented staff member, the committee can allocate the money evenly to staff or choose to differentiate bonuses in any way they decide. For example, some might choose to give all staff the same bonus amount (no variation); others might differentiate amounts based on a staff member's position, with teachers, for example, receiving one amount and secretarial staff receiving another amount (differentiation by job classification). Still others might choose to differentiate amounts within position type, for instance, varying teacher amounts based on a predetermined calculation, such as a teacher's influence on student performance. Note that these plans include all schools participating in the program, not just those that ultimately won the bonus.

### Overall, CC Members—Especially UFT Committee Members—Voiced a Preference for an Equal Bonus Distribution Plan

More than one-half of all CC members reported that (1) they believe all UFT members should receive an equal share because all make important contributions to school success, (2) most of their teacher colleagues had expressed similar preferences for an equal distribution plan, and/or (3) they were reluctant to consider anything but distributing equal bonus amounts to all employees for fear that an unequal distribution would negatively affect school climate (Table 5.3). UFT members were in fact significantly more likely to agree with two of the three statements than their administrative counterparts. As one teacher CC member reported on the open-ended portion of the 2009 survey, "everyone works hard together and succeeds or fails together and should share in the reward equally."

Nevertheless, even though the majority of CC members believed in equal distribution, they seemed to make one exception to this belief: More than one-half agreed that staff receiving unsatisfactory evaluation

**Table 5.3**
**Percentage of CC Members Agreeing or Strongly Agreeing with Statements About Equal Distribution Plan (2009–2010)**

| Statement | All Members | Admin | UFT |
|---|---|---|---|
| I believe all UFT members should receive an equal share of the bonus because all staff make important contributions to the success of the school | 63 | **50** | **72** |
| Most teachers in our school have communicated to me or members of the Compensation Committee that they are opposed to anything but an equal share distribution of the bonus award money | 61 | 58 | 63 |
| I am reluctant to consider anything but distributing an equal share of the bonus to all employees for fear that unequal distribution will negatively affect school climate | 59 | **47** | **66** |

NOTES: Statistically significant differences between administrators and UFT staff are indicated in boldface ($p < 0.05$). Total numbers and percentage calculations for each item exclude those who responded "don't know." See Appendix A for more detailed results on these items.

ratings deserved a smaller share of the bonus pool than others received. There was far less support overall for differentiating the amount of bonus awards based on other potential program criteria, including the extent to which staff demonstrated exceptional performance, provided direct academic instruction, had assignments in tested grades and subjects, and served as mentors or school leaders.

However, as shown in Table 5.4, administrators and UFT members held disparate views regarding a differentiated distribution plan. For example, more than three-fourths of administrators agreed that U-rated teachers should receive smaller bonuses, while slightly more than one-third of UFT members agreed. In general, administrators on CCs were much more inclined to support differentiated bonus award

**Table 5.4**
**Percentage of CC Members Agreeing or Strongly Agreeing with Statements About a Differentiated Distribution Plan (2009–2010)**

| Statement | All Members | Admin | UFT |
|---|---|---|---|
| I believe staff with unsatisfactory evaluation ratings should receive less than others | 55 | **79** | **38** |
| I believe staff with exceptional performance should receive a larger share of the bonus than staff with lower levels of performance | 34 | **57** | **25** |
| I believe staff who provide direct academic instruction to students (e.g., classroom teachers) should receive a greater share than other staff with indirect and non–academic support to students (e.g., counselors, nurses) | 32 | **41** | **26** |
| I believe staff who do extra work at the school or work additional hours (e.g., tutor after school) deserve more than others | 32 | **50** | **21** |
| I believe teachers in subjects and grades that impact Progress Report scores deserve more than teachers in the other grades and subjects | 18 | **25** | **13** |
| I believe senior teachers who serve as mentors or school leaders deserve more than others | 13 | **19** | **9** |

NOTES: Statistically significant differences between administrators and UFT staff are indicated in boldface ($p < 0.05$). Total numbers and percentage calculations for each item exclude those who responded "don't know." See Appendix A for more detailed results on these items.

amounts for staff than were their UFT colleagues. In fact, almost three-fourths or more of UFT committee members disagreed or strongly disagreed with these statements about differentiating bonuses to staff (excluding the question about U-rated staff).

### CC Preferences for Egalitarian Awards Reflected the Opinions of Many but not All Staff Members in the Case-Study Schools

Consistent with the survey results and distribution plan data, many individuals we interviewed in the case-study schools expressed strong beliefs that all members of a school community contributed to student success and therefore deserved an equal share of the SPBP bonus. "We didn't want anyone to feel excluded," explained one UFT committee member, "We feel everyone works really hard here, from the secretary to the paraprofessionals. Everyone's position was equally valued at the school." Nevertheless, the committee in this school awarded smaller bonuses to nonteaching staff. Many case-study respondents felt just as strongly that an egalitarian distribution would contribute to greater cohesion and collaboration. Some interviewees also believed that no objective measures were available to "fairly" differentiate.

However, some interviewees were open to the idea of differentiation and rewarding certain staff more for greater contributions or performance but did not attempt to do so, citing a fear that it would generate divisiveness among staff. This group of individuals viewed the egalitarian plan as "easier" and more palatable. "I don't want people to get upset," admitted one teacher who liked the idea of giving larger shares to tested-grade teachers but ultimately decided to support an egalitarian plan. One UFT chapter leader believed that deciding which staff members deserved more would be "destructive to the fabric you hope to create" in a school. A colleague in the same school concurred:

> If you say certain teachers should get more, then I think you have jealousy and lack of collaboration which is a big-time part of teaching . . . When money starts to be the barometer of performance, then I think you are going down a slippery slope of isolation and alienation among faculty.

NYCDOE and UFT leaders were not surprised that most schools did not differentiate bonus award payouts. As one NYCDOE leader explained, it would take a "big cultural leap to move in that direction."

Nevertheless, a minority of individuals interviewed in case-study schools favored differentiation and the idea of rewarding staff who go "above and beyond" (a phrase we heard repeated across schools). These individuals were often younger staff, career-changers, or principals. One elementary school principal explained,

> I think the grade 3, 4, 5 teachers should get a little bit more. Another thing I said that if anyone got a U-rating, you're not getting the bonus. Because then a wrong message is being sent. My attitude is that you get a paycheck twice a month. You're here to do a job—you get paid to do a job. And you should look at the bonus as okay if you go above and beyond, then this is something that's nice that somebody gives us.

Similarly, a teacher in a tested grade level at another elementary school felt strongly that schools should not reward "bad workers" or else everyone will "slack off." This teacher implied that the system is not fair if it rewards individuals who do not work hard just the same as individuals, like herself, who come to school early to prepare, who stay in close contact with parents, and who continually develop new instructional materials.[3]

### Teacher Survey Respondents also Supported Equal Awards but Mostly for Teachers

As Table 5.5 indicates, almost two-thirds of teachers indicated a preference for distributing equal shares of the bonus to all school staff. Yet, almost the same proportion of teachers also reported that non-teachers should receive a smaller share than teachers. Only a minority

[3] Interestingly, on both ends of the egalitarian-to-differentiated spectrum, individuals defended their positions as being fair. One high school counselor defended an egalitarian plan by noting, "It was my understanding that it would be distributed evenly so that it was completely democratic, and no particular department or person would be eligible and more so that everyone in the building would work together as a team. It's the most fair way—it's very hard to figure out if you're going to pay a person more."

**Table 5.5**
**Percentage of Teachers Agreeing or Strongly Agreeing with Statements About Bonus Distribution Plans (2009–2010)**

| Statement | Percentage |
|---|---|
| All building personnel should receive an equal share of the bonus | 64 |
| Non–teaching staff should receive a smaller share of the bonus than teachers | 62 |
| Non–teaching staff should not be eligible for a share of the bonus | 26 |

NOTES: See Appendix B for more detailed results on these items.

(26 percent) believed that nonteaching staff should not be eligible for a bonus share.

### The Distribution Plans Participating Schools Submitted Reflected a Strong Preference for Egalitarian Award Distributions

Using data from the NYCDOE Galaxy system, we studied the distribution of bonuses slated for every UFT staff member in every SPBP participating school if the school earned the bonus. Across all 196 schools that participated in SPBP in 2010, 82 percent of all staff were slated to receive the modal or most common award for their school. The modal award ranged from $2,294 to $4,500 with a mean of $3,068; in 63 schools (32 percent), the modal or most common award was $3,000, and in 179 schools (91 percent), it was within $500 of $3,000. Consequently, the most common award amount was $3,000 in 2010, with 29 percent of staff targeted to receive that amount. This was also the most common award in both prior years: 52 percent of staff were allocated $3,000 in 2008, and 38 percent of staff were allocated $3,000 in 2009.[4]

[4] The greater proportion of proposed awards equaling exactly $3,000 in 2008 and 2009 corresponds directly to a greater percentage of schools setting the modal award for the school to $3,000 (57, 41, and 32 percent for 2008, 2009, and 2010 respectively). It appears that many CCs balanced awards so that the most staff received exactly $3,000 in 2008. In later years, CC allowed the mode to go above or below $3,000 (more often above) rather than balancing awards that were not equal to the mode so that the mode could remain exactly $3,000.

Even though the vast majority of staff were targeted for a single award in each school, there was a substantial range in awards from $0 to $6,628 in 2010, which was somewhat greater than in the first two years of the program ($0 to $5,914 in 2008 and from $0 to $6,200 in 2009). Adjusting for the number of staff within each school, the within-school average bonus amount proposed was $2,856.65 (SD = $183.24) in 2008, $2,840.88 in 2009 (SD = $191.61) and $2,811.91 (SD = $186.95) in 2010. All the participating schools across all three years chose to disseminate bonus money to both teachers and noninstructional staff (e.g., paraprofessional, administrative assistant).[5]

Among staff members who were targeted not to receive the modal award, 86 percent were to receive less than the modal amount on aver-

[5] NYCDOE determined the overall total award based on each school's table of organization (TO), which is a snapshot of the payroll on October 31, listing all staff employed at the school, sorted by UFT title and whether he or she was full or part-time. Participating schools were given this list and allowed to add any additional UFT-represented staff members they believed deserve a share of the bonus or remove names of individuals not deserving of a share. For example, CC members may decide to add to the list staff members who started working at the school midyear or remove individuals who left early in the year. Guidance from NYCDOE and UFT at information sessions to CC members in 2010 explained:

> There may be employees on the TO who are not working at the school this year (sabbaticals, leaves, etc.) and the CC is under no obligation to award a bonus to them, though it can do so. The school receives $3,000 for the presence of these employees on the TO but the CC can distribute the available pool as it sees fit. (NYCDOE and UFT, 2010)

The guidance also reminded participants that the initial agreement between UFT and the NYCDOE setting up the program states: "the Compensation Committee shall presume that all UFT-represented staff employed at a school that meets the targets for the bonus have contributed to the school's achievement to some extent and therefore should share in the bonus" (NYCDOE and UFT, 2010). A school received $3,000 for every UFT member on the TO list. The Galaxy data included UFT members on the TO list and every member the school added—the universe of all members who were potentially eligible for an award if the school won the bonus. Hence, the total number of potentially eligible UFT members in a school could exceed the number on the TO list used to determine the total bonus to the school. Consequently, the average award for members in the Galaxy data or for all potentially eligible members was often less than $3,000 per school. For example, suppose a school had 40 eligible members on the TO list, the CC for the school would have $120,000 = 40 x $3,000 to allocate. Now, also suppose the school added two members who joined its staff after October 31. The total potentially eligible UFT staff is 42, so the average award per total eligible staff is $120,000/42 = $2,857.14.

age across school plans. Overall, 82 percent of distribution plans that included any deviations in award from the modal amount had only negative deviations. For the most part, the distribution plans schools used called for awarding a modest percentage of staff smaller awards than the per-staff total award and then readjusting the proposed awards for the remaining staff to receive equal allotments of the remaining total dollars. For instance a plan might identify one of the school's 40 staff members to receive one-half the award ($1,500) and distribute the remaining $1,500 evenly among the remaining 39 staff members, giving each an award of $3,038.

The tendency for most school distribution plans to award nearly all teachers the modal award also held in the first two years of the project. In fact, the percentages of staff slated to receive the modal award were significantly higher statistically in the first two years of the program than they were in 2010. In 2008, plans targeted 86 percent of staff on average to receive their schools' modal awards; in 2009, the average rate was 85 percent.

The percentage of staff a school's plan targeted to receive the modal award correlated across years. When plans called for relatively small percentages of staff to receive the modal award in one year, rates of staff receiving modal awards in the next year or even two years later were somewhat more likely to be lower. However, the correlation was generally weak and weaker between 2008 and 2010 than between adjacent years. Hence, it is not the case that schools developed plans that would result in significant variation in awards among their staffs and then consistently implemented these plans across years.

### As Reported in the Case Studies, Most CCs Did Not Develop Completely Egalitarian Plans, but Most Avoided Judging Performance

Only 14 schools (7 percent) did not differentiate awards in some way among staff. The remaining schools had some differentiation in the awards for eligible staff. This was a decrease from the priors when 9 percent did not differentiate in 2009 and 18 in 2008.

However, differentiation in most schools was limited to a small percentage of the staff, and the factors used to differentiate proposed

awards tended to be unrelated to individual performance and outside the individual staff member's control. Data collected by the NYCDOE Galaxy system four-question survey on the design of the distribution plans provided information on the specific factors CCs used to determine awards for their staffs in year 3.[6] The survey distinguished two types of differentiation of awards: excluding a staff member from the award pool (i.e., slating a staff member to receive no award) and differentiating among members who shared in the award pool. Seventy-six schools reported that at least some staff would be slated to receive no award.[7] By far, the most common factor used as the basis for this determination was whether or not an individual completed the full year at the school (Table 5.6). Consistent with our case-study findings, 24 percent of the 76 schools (18 schools) reported slating U-rated staff to receive no award. Attendance was the next most common factor (used by 7 percent of the 76 schools), and only one school reported using demonstrated low quality on the basis of student performance as a factor for not providing a bonus to staff.

---

[6] The four survey items on the NYCDOE Galaxy survey were:

(1) In the distribution plan developed for your school this year (2009–2010), if your school meets its Progress Report performance target, would any full-time UFT staff members receive $0 or no share of the bonus award pool? (Select One); (2) On what basis would individuals receive no share of the bonus award pool? An individual would receive no share if he/she . . . (Select All that Apply); (3) In the distribution plan developed for your school this year (2009–2010), if your school meets its Progress Report performance target, would every employee put on the list—excluding those who received $0—receive approximately the same share of the bonus award pool (e.g., everyone receiving about $3000)? (Select One); (4) On what basis would individuals receive more versus less of the bonus award pool? The plan awards more or less of the bonus based on . . . (Select All that Apply).

Only respondents reporting staff could receive an award of $0 on Question 1 were asked Question 2.

[7] Three schools did not complete the Galaxy survey, and 177 reported that they either gave some staff awards of zero or differentiated awards. Nine of the 16 schools that did not report differentiating awards on the survey actually showed differentiation in their proposed awards. Similarly, in six of the schools that did not differentiate awards indicated that they did in their survey responses. This could indicate errors in reporting or that respondents reported factors that would be used to set awards, if applicable, even if some were not applicable. For instance, a respondent might report that a U-rating would be used to withhold an award, if that were CC policy, even if no staff members actually had a U-rating.

**Table 5.6**
**Percentage of Schools Withholding Awards from One or More Staff Members That Report Using Criterion to Decide Which Individuals Receive no Share of the Bonus Award Pool**

| Factor | Percentage of 76 Schools |
|---|---|
| Does not complete the full year at the school (e.g., a midyear entrant/departure) | 79 |
| Is a U–rated staff member | 24 |
| Has a part–time assignment at the school (e.g., split between multiple schools) | 14 |
| Has low attendance during the school year | 7 |
| Demonstrates a lower quality of performance based on other student achievement results (e.g., periodic assessment results) | 1 |
| Member refused | 1 |
| Member had charges pending | 1 |

NOTE: An additional 2 percent of schools reported not knowing the reason.

Schools also reported the factors they used to determine the amount of the awards allocated to staff who were scheduled to receive a share of the bonus. Again, time in the school, in terms of working full time or part time or full year or part year, was by far the most common set of factors. About one-sixth of schools used individuals' attendance as a factor. Less than 10 to 15 percent of schools reported relying on evaluations of staff performance or job titles, and even fewer relied on type of assignment. Twenty schools (just over 10 percent) reported factoring in seniority, which the formal SPBP explicitly prohibited. Only 7 percent of schools (13 schools) reported using U-ratings to determine awards. Finally, just 4 percent (8 schools) and 3 percent (6 schools) of CCs reported performance-based value-added or other student achievement, respectively, as a factor. Table 5.7 summarizes the percentages of schools using various criteria for differentiation.

**Table 5.7**
**Percentage of All Schools Reporting Using Specific Criteria to Differentiate Bonus Shares**

| Factor | Percentage of All Schools |
|---|---|
| Whether individuals had a full–time or part–time assignment at the school | 49 |
| Whether individuals completed the full Year at the school compared to midyear entrants or departures | 41 |
| Individual attendance during the school year | 17 |
| Other means or means unknown | 16 |
| Whether individuals had a job title that involved direct instructional work with students | 14 |
| Whether individuals had a job title of teacher or classroom teacher | 10 |
| Individual length of service at the school or in the NYC school system | 10 |
| Designation as U–rated staff member | 7 |
| Individual quality of performance based on an evaluation of practice, such as formal evaluation by supervisor | 5 |
| Hours individuals devoted to school activities (e.g., clubs, lunch duty) or additional responsibilities (e.g., team leader grant writing) | 5 |
| Whether individuals had an assignment related to tested subjects and grades (e.g., ELA teacher, math coach) | 4 |
| Individual quality of performance based on value–added assessment results | 4 |
| The average salary for their job title (e.g., everyone received a fixed percentage of the average salary for their job title) | 3 |
| Individual quality of performance based on other student achievement (e.g., periodic assessment results) | 3 |
| Whether individuals had an assignment related to high-needs students (e.g., special education, ELL) | 1 |
| Pro–rated F–status teachers[a] | 1 |

[a] A per diem full- or part-time employee engaged for a full term but for less than five full days per week (e.g., for 2 days a week).

### Case-Study Schools Provide Insights into How Factors Were Used to Determine Bonus Awards Among Staff

To understand better how and why committees differentiated bonuses among staff, one of our sampling criteria for case-study schools was the degree of differentiation in their plans, with oversampling of schools that differentiated substantially. Of the seven case-study schools we visited in 2009, five differentiated the bonus to some degree in year 1. In 2010, five case-study schools had some level of differentiation of bonus payout within their plans in one or multiple years of the program. In some cases, differentiation was based mainly on job title and/or subject taught. For example, two of the 14 schools allocated smaller bonus shares to some or all paraprofessionals, secretaries, guidance counselors, and social workers so that more could be awarded to some or all teachers. In one school, the extra money generated from giving a smaller portion to each paraprofessional and secretary was used to fund bonuses for substitute teachers who had spent most of the year in the school covering for absent teachers. Two other schools differentiated shares within the group of classroom teachers, giving larger bonuses to staff viewed as driving test scores and responsible for test scores, which factored into Progress Report targets. One middle school allocated more to teachers of "core content" subjects (ELA and mathematics), while another school gave more to upper elementary grades.

In contrast, five of the schools we visited chose to differentiate at least some of the bonus shares based on performance or perceived merit rather than job title or role. One school provided all staff with a guaranteed equal share of the bonus pool and then allocated additional funds to staff who demonstrated that they had devoted time to at least five specific types of school activities or additional responsibilities, such as overseeing lunch detention, writing grants, and leading a professional development session. Maintaining 90 percent attendance and 0 percent "lateness" was another factor that could qualify in the list of five. Thus, performance was based on demonstrated extra effort or time.

In four other schools, decisions about how to differentiate the bonus were based on committee discussions about who they felt contributed more or less to the school's success rather than explicit objec-

tive criteria. In one of these schools, CC members described basing these decisions on attendance in schools, principal and teacher-led observations, and general perceptions of individuals' efforts or willingness to go "above and beyond." Some acknowledged that the decision was not always based on objective criteria, noting that "one person got whacked by the principal for pissing him off." In other schools, CC members could not provide a concrete description of criteria used to award some staff more or less of a share but instead explained that "we know the ones who participate." As described further below, this lack of objective criteria created some anger and resistance among staff. As a result, one of the principals we interviewed in year 2 backed away from supporting a merit-based plan in year 2, and another principal made some adjustments to the process in year 2, such as asking staff to nominate individuals they felt deserved larger bonuses.

Across time, the majority of case-study schools made only minor changes to their distribution plans. Although many schools adjusted aspects of the plan—for example, adding a few individuals to the list of recipients, prorating for time spent in the building, or giving less or $0 to U-rated staff—schools generally did not make wholesale changes in their overall approach to distributing funds. In many respects, the egalitarian or differentiated approaches became highly institutionalized within the school, and staff appeared unwilling to and uninterested in considering major changes over time. As discussed below, attempts to move toward more differentiated approaches generated significant backlash in some cases. We observed a few exceptions to this pattern of stability over time. For example, one school adopted a highly differentiated distribution plan in year 3 after two years of having an egalitarian plan. In this school, the principal—in violation of SPBP rules—appointed the other three members to the CC, two of whom were new to the school.

### Most Distribution Plans Did Not Use Individual Performance When Determining Awards, but Those That Did Had Greater Variability in Awards

The CC member survey responses suggest that committees took three somewhat distinct approaches to developing plans: (1) a completely

egalitarian approach, in which all staff members, regardless of individual effort or performance or percentage of time worked at the school, received exactly the same amount; (2) a differentiated approach, with amounts based in part on individual performance; and (3) a differentiated approach, with amounts based only on job titles, the proportion of the school year the staff member had worked in the school, or whether he or she had worked at the school full time. All three methods appear to have been chosen in part based on the different opinions about what was fair and what type of performance was appropriate to reward.

Overall, we identified 14 schools with a completely egalitarian approach to awards in 2010. Survey responses indicate that, of the remaining 182 schools, 60 (31 percent) rewarded staff partly on the basis of individual performance, and 122 (69 percent) differentiated awards only on the basis of factors other than individual performance.[8] Schools were classified as differentiating on the basis of individual performance if they reported they had provided no awards to staff with low attendance, a U-rating, or lower quality of performance as measured by value-added assessment results or if they reported determining nonzero award amounts on the basis of individual attendance; quality of performance based an evaluation of practices, such as formal evaluation by supervisor; other student achievement results; or value-added, or hours individuals devoted to school activities (e.g., clubs, lunch duty) or additional responsibilities (e.g., team leader grant writing). Eighteen schools reported plans that would reward no bonuses to U-rated staff; only one school reported withholding bonuses on the basis of value-added. Attendance was the most common individual performance

[8] In all, 193 schools responded to the Galaxy four-item survey. The schools that did not respond were scored as not having differentiated using individual performance because data on parallel items from the RAND CC member survey for the three schools indicated that they did not use individual performance for determining award amounts. Responses to the items on the RAND CC member survey that were parallel to the Galaxy items suggested that four additional schools might have differentiated awards using individual performance. The results reported here, other than the number of schools in each group, remained virtually unchanged when we included these schools in the group that differentiated on the basis of individual performance. However, inconsistencies in the CC member responses led us to believe the Galaxy data were more reliable. Hence, we report results based only on the Galaxy data.

indicator used to determine award size (33 schools). Fourteen schools reported using U-ratings to determine award sizes for staff receiving awards, and only four of these schools were among the 18 that used U-ratings to withhold bonuses. Ten schools reported using quality of performance based on an evaluation of practices to determine awards; nine reported using hours devoted to school activities; eight reported using quality of performance based on value-added; and six reported quality of performance based on other student achievement. Sixty percent of schools reported using only one criterion to determine awards, but the remaining schools combined criteria, most often combining attendance (13 schools) or U-ratings (10 schools) with other measures. For 26 of these schools, attendance was the only personal performance measure used in determining awards. The remaining 34 schools (17 percent of all schools) used at least one other performance indictor in determining awards.

Compared with schools that did not consider individual performance to determine awards, distribution plans from schools that used individual performance as at least one factor for determining awards were less likely to award staff the modal award for their school. On average, schools using individual performance to determine bonus awards targeted 74 percent of their staffs to receive the modal awards; on average, other schools that differentiated awards targeted 82 percent of their staffs to receive the modal award. This 8-percentage-point difference is statistically significant. Distribution plans from schools that used individual performance for determining awards called for about 4 percent of their staffs on average to receive zero dollars. In contrast, plans from other schools that differentiated awards awarded zero dollars to only about 2 percent of their staffs on average. Again, this difference between schools is statistically significant. However, the average amount that awards deviated from the modal award differed only by about $103 between these two groups of schools, and this difference was not significant. In sum, the distribution plans of schools that used individual performance in the determination of awards were much more likely to differentiate awards than other schools.

The Gini Coefficient is commonly used to quantify the extent of deviation from a uniform disbursement (Gini, 1912). A low value for

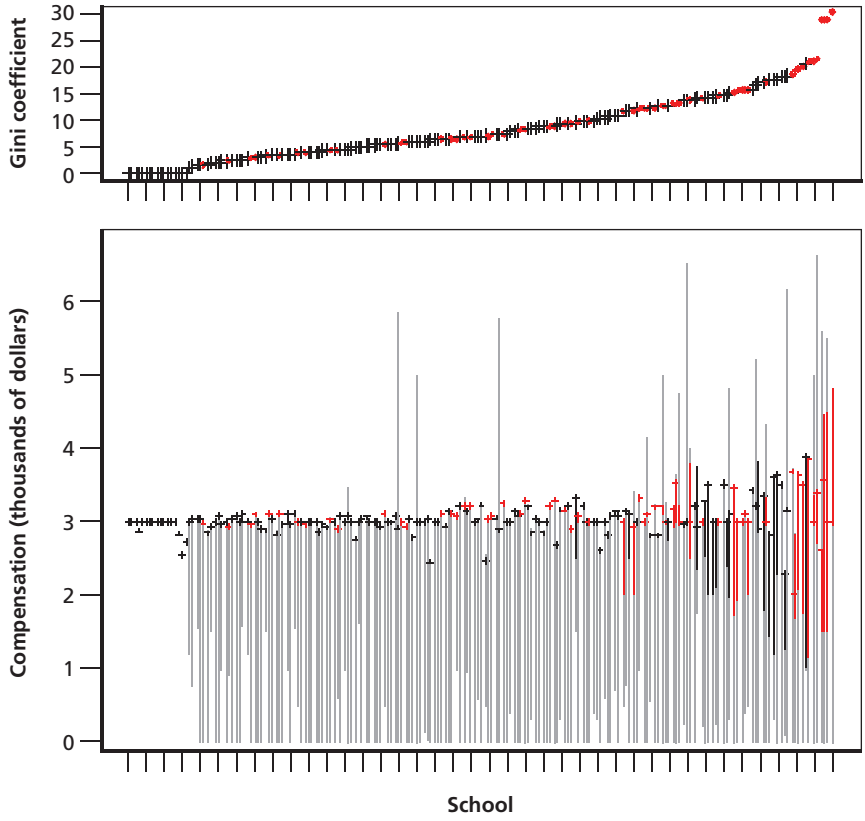
the Gini coefficient indicates a more egalitarian award distribution plan (0 denotes equal awards to all staff), and a larger value indicates a more unequal distribution (100 indicates perfect inequality). For instance, one award distribution with a Gini Coefficient of 1 awarded for 89 of 91 staff $3,000 and the remaining two staff (both paraprofessionals) $1,500. An example of a plan with a Gini Coefficient of 5 awarded $3,000 to 56 of 59 staff and zero for one teacher and two administrative support staff. Even Gini Coefficients of 20 correspond to only mild deviations from uniform allocation. For instance, in one plan with a Gini Coefficient of 20, 29 of 50 staff, including all five paraprofessionals, one of two administrative assistants, and one of three administrative support personnel, received $2,000; 20 of the staff, including 18 of 21 teachers, received $4,500; and one teacher was allocated nothing.

The Gini coefficients of all schools were generally very small because very large percentages of staff were targeted to receive the modal awards at the schools. However, the distributions of awards in schools that used individual performance in the determination of awards were less egalitarian than other schools. The average Gini Coefficient for schools that used individual performance in determining awards was 12.0, while it was only 8.0 for other schools that had some deviation in awards.

Figure 5.2 highlights the differences among the groups and shows the general distribution of planned awards for all SPBP schools in 2010. The top panel displays each school's Gini Coefficient. Schools that used individual performance in the determination of awards are denoted by red dots, and other schools are denoted by black plus signs. The vast majority of values were very small, indicating limited deviation from uniform allocation of bonus awards. Schools that used individual performance in determining the awards were spread across the range of Gini Coefficients but were concentrated at the top of the distribution, indicating that the awards were more variable and less uniform at these schools.

The bottom panel of Figure 5.2 summarizes the award distribution plan for each school participating in SPBP. The figure plots the range of awards as a vertical light gray line, with the upper end of the line at the maximum award and the lower end at the minimum award.

**Figure 5.2**
**Gini Coefficients and Award Distributions by SPBP Schools for the 2010 School Year (schools using individual performance in determining awards highlighted in red)**




RAND *MG1114-5.2*

For each school, the range from the 25th to 75th percentile in awards is plotted as the vertical black or red line (25th percentile at the lower end of the line and 75th at the upper end; red lines are used for schools that used individual performance in determining awards and black lines are used for all other schools), and the median award is plotted using a plus sign. As shown in the figure, most schools do not have black or red vertical lines, indicating that a very large majority of the schools participating in SPBP had no variation in bonus awards between the

25th and 75th percentile. For these schools, at least 50 percent of the staff would receive the same bonus amount if the school were to earn a bonus; in only handful of schools (those with Gini Coefficients over 20) was the variation between the 25th and 75th percentiles sizeable. Figures for the early years are quite similar (see Appendix C).

The 26 CCs that used performance factors other than attendance when determining awards were even more likely to differentiate awards than the other CCs for which attendance was the only personal performance measure used in determining awards. On average, only 70 percent of staff received the modal award for these schools, compared with 80 percent for schools that did not use factors other than attendance.

### Award Distribution Was Not Related to Whether the School Had Won the Bonus in the Past, but Award Distribution Was More Egalitarian in Larger Schools

Neither the change in the level of differentiation that the Gini coefficient measures nor the percentage of staff not receiving the modal award for the school is related to whether or not a school won a bonus for the previous year's performance. Schools decreased the proportion of staff slated to receive the modal award for their schools each year, but the changes were similar for schools that did and did not earn a bonus in the previous year. Similarly, the Gini coefficients increase or essentially stay the same across years, and the changes are very similar for the schools that did or did not earn a bonus the previous year. Thus, it does not appear, on average, that schools used differentiation in response to their performance in the previous year.

Similarly, reported use of differentiation did not predict whether or not a school would win a bonus in the current or next year (even after controlling for the schools prior year bonus status). This is true for the level of differentiation the Gini coefficient measures, for the percentage of staff not receiving the modal award, and for whether or not the CC reported using personal performance (with or without attendance) to determine awards.

Either the level of differentiation the Gini coefficient measures or the percentage of staff not receiving the modal award for the school is, however, related to school size as measured by the number of eligible

staff. In general, larger schools (those with more eligible staff) had significantly less differentiation of awards. Overall, size was unrelated to whether or not a school differentiated, but among schools that differentiated awards, there was a significant negative relationship between the Gini coefficient and size every year and between the percentage of staff not receiving the modal award and size in 2009 and 2010, with a negative but nonsignificant relationship in 2008. The relationships were strongest in 2009. For instance, among schools that differentiated awards, the average percentage of staff receiving the modal awards for schools below median size in 2010 was 23 percent, but it was only 17 percent for schools above median size that year. In 2009, the numbers were 22 and 14 percent. Thus, although one may expect larger schools to have a higher potential for "free riders" and a greater incentive to differentiate awards by individual performance to motivate all individuals to improve performance, our findings do not support this supposition.

## Response to Distribution Plans

This section examines staff responses to the bonus distributions—both the plans and the actual payouts that occurred in schools earning the bonus.

### Distribution of Bonus Shares Went Smoothly in Most Schools; Only a Minority of Respondents Reported Distribution Problems in Their Schools

Overall, according to the majority of respondents in 2010, the distribution of bonus shares appeared to go very smoothly and was perceived to be quite fair. Nevertheless, a significant minority of teachers reported problems. For example, 18 percent of teachers reported that their CCs had distributed the bonus money unfairly. As Table 5.8 illustrates, less than one-third of CC respondents reported that colleagues had been left off the list of individuals receiving checks or that colleagues were upset about the unfair distribution of the bonus. These numbers were slightly higher in 2009 (although differences are not statistically sig-

**Table 5.8**
**Percentage of CC Members Agreeing or Strongly Agreeing with Statements About Bonus Distribution (2009–2010)**

| Statement | All Members | Admin | UFT |
|---|---|---|---|
| Some staff at our school chose to donate a portion of their bonuses to someone who was otherwise ineligible (e.g., non–UFT represented staff, volunteer, security agent) | 28 | 23 | 31 |
| Some staff members at our school were upset about the unfair distribution of bonus award payments | 23 | 21 | 24 |
| Some staff members at our school who should have received a share of the bonus award pool were left off of the list of individuals receiving checks | 17 | **12** | **20** |
| Some staff members at our school chose to donate their bonus awards to the school | 6 | 5 | 6 |

NOTES: Statistically significant differences between Administrators and UFT staff are indicated in boldface ($p < 0.05$). Total numbers and percentage calculations for each item exclude those who responded "don't know." See Appendix A for more detailed results on these items.

nificant), suggesting in particular that CC members were more aware of the need to check the official Table of Organization sent to them by NYCDOE and add staff who might have been missing from this list.

In case-study schools, CC members reported taking great care to account for all staff contributing to their school's performance, including long-term substitutes and other individuals who had been left out in previous years. Only one school visited in 2010 reported problems with distribution implementation in year 2. In this school, CC members required staff to apply for receipt of what was deemed the "incentive pool." Some staff were reported to have missed the application deadline and did not receive bonus shares despite qualifying for these funds and, in many cases, reporting that they were unaware of the deadline. In response to the fallout, in year 3, CC members changed the process to ensure this scenario would not repeat.

NYCDOE and UFT leaders reported relatively few problems with the distribution process. Neither NYCDOE nor UFT said they received many complaints each year, although they did hear from some

people who should have received bonus dollars who were omitted in both the first and second years of the program from the list the school submits to NYCDOE. "There were a fair number of people left out," said one official, "People were saying, 'I didn't get my bonus and I was at this school.' The school hadn't put them on the list. If this information does not come out until after the payout occurs. . . then there's nothing that can be done. There's no more money." Despite these problems, none reached the level of an official appeal to the oversight committee.

Interestingly, more than one-quarter of survey respondents reported that some staff at their school donated a portion of their bonus to individuals who were otherwise ineligible to receive a share, such as non-UFT-represented staff members and volunteers. Fewer reported that some staff donated their bonuses to the school.

### Differentiated Distribution Plans Were Seen as More Risky Than Egalitarian Plans and, Once Implemented, Generated a Significant Backlash in Several Case-Study Schools

In a few case-study schools that pursued differentiation, there was evidence that those fearing negative consequences were accurate in their predictions. Several schools faced fallout from staff members after they discovered that some individuals received a greater share of the bonus than others. In one school, a small group of individuals expressed dissatisfaction with the amounts they had received relative to others, and one tried to file a formal grievance (later discovering it was not possible). In two other schools, this conflict centered on the lack of transparent, objective criteria for differentiating bonus shares. In both schools, staff felt they had not received clear information about the basis for awarding some staff more than others and speculated over possible favoritism on the part of CC members. In turn, many questioned the fairness of SPBP. Some schools appeared to understand the potential for conflict resulting from a differentiated plan more than others. For example, one school declared that members of the CC were ineligible for a greater share of the bonus to avoid the perception of a conflict of interest. In contrast, in another school that did not take this into consideration,

several staff members who received a lesser amount questioned the fairness of the process in which committee members earned more.

A UFT leader's attempt in one school to engage the faculty in a discussion about the possibility of giving a greater share of the bonus to certain staff—using classroom teachers and tested-grade teachers as an example to spark conversation—created significant tension. Although the leader's intent was to push staff to think about and discuss the issue, the result, according to the principal, was that he "stirred up controversy." Many individuals reported that "it got crazy" and "unpleasant" with many colleagues feeling angry about the implication that they did not work as hard. One teacher reported that it took weeks to "calm people down." For another teacher, the experience raised the question of "Is this program worth it? Enough to make it an unpleasant place to work?" Eventually, the UFT leader (who had been voted to participate on the CC) stepped off of the committee and called for a revote to avoid escalating tensions further.

### The Bonus Distribution Plan Was Opaque in Many Case-Study Schools, Particularly in Those That Differentiated Award Amounts

Another indicator of the highly charged nature of differentiated plans is the fact that many CCs and school leaders that created such plans purposefully limited communication about them. Some did not intend to share their plans with staff until NYCDOE publicly announced the bonuses and their schools had, in fact, received the bonus. Others intended to keep their plans secret even after the bonus announcement, in hopes that they could let the information "slide under the radar." These individuals believed that it was not worth unnecessarily "rocking the boat." One principal admitted, "I just feel like they [staff] don't need to know."

In fact, in recruiting schools to participate in our first year of case studies, we contacted the principal of a school that developed a highly differentiated distribution plan in year 1 who refused our request to visit, citing concerns that our visit would "stir things up." The principal explained that they had built a strong community this past year and implied that questions about why certain staff members received more funds than others might threaten that harmony.

We interviewed many individuals in schools with differentiated plans that had won the bonus who either did not know whether they had received more or less than colleagues or, if they were aware of receiving more or less, did not know why. When asked why some individuals received more or less, one paraprofessional replied, "I assume it is because whoever is on the committee felt that they didn't do their job as well that year." In some cases, the lack of transparency created discomfort for individuals who knew they had received more than a friend or knew why a colleague had received less but did not want to give them this information or felt awkward possessing such privileged information. The lack of transparency also led some to question the motives or ethics of CC members. For example, some individuals questioned why certain staff serving on the CC earned larger shares than those not on the CC. In other cases, we were told of rumors about individuals perceived to be "the principal's pet" receiving more than other staff. A few individuals identified the lack of transparency as a problem and recommended that there be full disclosure of distribution plans. At least one school that experienced some staff fallout over the lack of transparency made a conscious effort in year 3 to publicize the distribution plan before the end of the year.

Even in schools that did not have highly differentiated plans, there were many instances of CC members making minor adjustments to dollar amounts or adding individuals to the list of recipients but never sharing the rationale for these decisions publicly.

## Summary and Discussion

Our data suggest that the CC process was implemented fairly smoothly, although there is evidence that some schools encountered some difficulty in the decisionmaking process. More specifically,

- Schools formed their four-person committees and generally followed guidelines regarding membership and procedures (although not all UFT CC members were formally elected).

- Most CC members reported that the decisionmaking process was fair, collegial, and inclusive and that reaching consensus was easy—although some other survey and case-study data suggest otherwise. For example, 44 percent of teachers disagreed or strongly disagreed that teachers' preferences were taken into account when developing the distribution plan. Also, some CC members questioned whether the requirement that UFT-represented staff make up one-half of the members truly guaranteed an even playing field and spoke about power differentials that played out between administrators and UFT members in committee deliberations.

Consistent with strong preferences for egalitarian award distributions, schools tended to keep awards as uniform as possible in their distribution plans. In particular,

- CC members—particularly UFT-represented members—voiced a preference for an equal distribution plan. More than one-half reported that all school personnel should receive an equal share of the bonus and that they were reluctant to consider anything but equal distribution for fear that unequal distribution would negatively affect school climate. Administrators on the CC were significantly more likely to support differentiating bonus awards than their UFT counterparts.
- Almost two-thirds of teachers indicated a preference for equal distribution, yet almost the same proportion also reported that nonteachers should receive a smaller share than teachers.
- The most common individual bonus amount awarded to staff in the plans was $3,000 in all three years.
- Differentiation was limited in all three years. Most staff received the same award amount. In fact, 82 percent of staff were slated to receive the most common or modal award for their school.
- Differentiation tended to involve a few individuals being allocated less than others, which is consistent with the survey and case-study reports, which suggested that such factors as midyear

entrance and part-time status at the school were the primary factors for differentiation.

- Inconsistent with the notion that larger schools might use differentiation to offset free riders, larger schools actually tended to be less likely to differentiate awards.
- Among the schools that did not provide equal shares to all staff, the most common factors used as the basis for differentiation were whether an individual had a full-time assignment at the school or completed the full year at the school.
- About 31 percent of schools considered individual performance, such as absences, U-ratings, or other performance ratings, when determining the bonus.
- Schools that considered individual performance when determining bonuses were more likely than others to award staff less than the modal awards and more likely to award staff no money. Schools that used individual performance in the bonus determination generally tended to have greater disparity in their awards, suggesting that, as the criteria considered for differentiation expanded, a greater number of staff met them.
- As supported by the case studies, schools that differentiated based on performance factors appear to have accepted differentiation as desirable, used a greater array of criteria for differentiating awards, and consequently gave more staff no awards and more staff something other than the modal award at the school.
- However, most of these schools remained cautious about deviating from egalitarian awards. On average, 74 percent of staff in these schools received the modal award for their schools, and in these schools, the large majority of staff typically received the modal award.

Although schools receiving the bonus reported no widespread problems, several case-study schools encountered difficulties when implementing differentiated distribution plans. In particular,

- The distribution process was reported to run fairly smoothly, and only a minority of CC members reported that staff were left off

the list of recipients or upset with a perceived unfair distribution. None of the reported problems reached the level of official appeal in any year of the program.

- In several case-study schools, there was fallout when staff members considered the decisions and criteria used to differentiate as being opaque and subjective. Some schools, particularly those with differentiated plans, also did not communicate the final distribution plans to staff.

Although some case-study schools reported that power dynamics were at play and that a significant minority of teachers doubted the extent to which final plans reflected teacher preference, the final distribution plans they submitted to NYCDOE clearly indicate that the egalitarian preferences that UFT-represented CC members and teachers embraced had won out. Thus, even though many administrators serving on the CC expressed a preference for differentiating bonuses based on teacher evaluations, performance, and effort, very few school plans ultimately differentiated, and few schools reported using these criteria to differentiate. These findings confirm the strong egalitarian norms known to characterize the teaching profession (Lipsky and Bacharach, 1983). They are also consistent with the experiences of the GEEG program in Texas, in which teachers in general designed egalitarian plans for their schools.

Even though many principals approved of differentiated plans, and even some UFT-represented staff indicated a willingness to differentiate based on U-ratings, the vast majority of CCs developed fairly egalitarian plans. It is possible that anticipated fallout resulting from differentiated plans—as some case-study schools experienced—may have dissuaded CC members from actively pushing for differentiation. Our interviews suggest this may also be driven in part by CC members wanting to avoid being seen as evaluating their peers' performance (one interviewee questioned whether such reviews were even ethical). It is also possible that the requirement for consensus further compelled those favoring differentiation to set aside their preferences to preserve the school's opportunity to receive the bonus. The CC process may thus have succeeded in ensuring a school-based decision for distribut-

ing bonus shares, but it did not produce differentiation based on individual performance as some leaders had hoped it would.

In the end, awards were set to a constant amount or determined by very objective measures that differentiated the amount of work or the task. The exceptions were U-ratings and attendance. Attendance is again a very quantified measure of time on task. A U-rating is a formal, well-established measure of very poor performance. CC might have viewed it as a safe, objective measure that most staff would accept as an indication that someone had not performed an equal job. Other performance measures are less well established and would have required creating rules correlating evaluation scores to a bonus amount—for instance, determining what was sufficiently poor performance to warrant no award. Consequently, use of such measures was rare.

Given the limited amount of differentiation instituted in the school distribution plans, the program generally did not include a mechanism to directly reward individual performance for the vast majority individuals. It is unclear whether this would have been necessary to motivate change. The theory is equivocal on this issue, but argues for increased collaboration and lack of competition with group awards. Empirical results are also mixed. Some research suggests that group awards can be more effective than individual awards (e.g., Condly et al., 2003), but international studies contradict this (see earlier literature review). Recent experiments find neither individual nor group awards yielded better student outcomes. Regardless, if program designers want some amount of individual differentiation of awards, SPBP CC experience suggests more guidance or stronger guidelines for the committee may be necessary. Differentiation did increase over time, so it is possible that time would yield more differentiation, but the rate of change was very small, and the base rate was very low. At the current pace, large-scale differentiation would be unlikely to occur for a very long time.

CHAPTER SIX

# Implementation of the Schoolwide Performance Bonus Program: Perceived Effects of the Bonus and Program Participation

Implicit in the SPBP's theory of action was a hypothesis that participating in the program would result in several intermediate outcomes that would ultimately improve student achievement. First, the program was expected to influence a school's improvement efforts and lead individuals to work together to implement strategies that would increase the probability of improving its Progress Report grade and winning a bonus. Under a schoolwide pay-for-performance program (and most accountability schemas), one would also expect schools to focus on what gets measured and used as criteria for winning the bonus, in this case, the key components that make up the Progress Report grade and target. Second, participation was expected to influence individuals by motivating staff members to improve their practices, collaborate more frequently, and/or remain at their then-current schools to help meet school goals.

This chapter examines the perceived effects of program participation on school improvement efforts and on individuals. While Chapter Seven presents direct estimates of program effects on student achievement, Chapter Eight presents a direct estimate of the effect of participating in SPBP on reported behaviors and attitudes through a comparison of teachers in SPBP and control schools. Perceptions of effects might not align with actual effects for many reasons, which we will discuss later. However, perceptions provide insights into how staff viewed the program and how they viewed their interactions with it. These data

can be valuable for interpreting the results in Chapters Seven and Eight and for determining how to modify a program to make it most effective at yielding its desired outcomes.

This chapter addresses the following research questions:

1. What perceptions did staff members have about SPBP's effects on school improvement efforts?
2. What perceptions did staff members have about SPBP's effects on individuals, including themselves, other colleagues, and students?
    a. What were the reported effects of winning or losing the bonus on practices and motivation?

## School Improvement

This section addresses staff reports on how SPBP participation influenced school improvement efforts and the extent to which it led individuals to work together to implement specific improvement strategies.

### More Than One-Half of CC Members Reported That Participation in the Program Increased Their Focus on a Variety of School Improvement Areas, Particularly Those Measured by and Factored into the Calculation of Progress Reports

Approximately two-thirds of all respondents reported that participating in SPBP had caused them to focus either slightly or significantly more on many areas of school improvement that contribute to the calculation of school Progress Reports, including the academic performance of high-needs students and of students on the cusp of state test levels (e.g., Level 2 and 3); communication of educational goals and opportunities for feedback; academic expectations; stakeholder engagement; student performance in ELA and mathematics (elementary, middle, and K–8 only); and in the case of high schools, student performance on Regents Exams, credits earned by students, and graduation rates (Table 6.1).

**Table 6.1**
**Percentage of CC Members Reporting Effect of SPBP on Focus of School Improvement Efforts (2009–2010)**

| | Did Not Affect Our Focus on This Area | | | Has Caused Us to Focus Significantly or Slightly More on This Area | | |
|---|---|---|---|---|---|---|
| **Focus of School Improvement** | **All Members** | **Admin** | **UFT** | **All Members** | **Admin** | **UFT** |
| Student attendance | 53 | 56 | 51 | 47 | 44 | 49 |
| School safety | 63 | 66 | 61 | 35 | 32 | 38 |
| Academic performance of ***high-needs*** students (e.g., ELL, SpEd, low–performers) | 38 | 41 | 36 | 61 | 59 | 62 |
| Academic performance of ***high-achieving*** students | 45 | 44 | 46 | 54 | 55 | 53 |
| Academic performance of students ***on the cusp*** of state test levels (e.g., between Levels 2 and 3) | 37 | 37 | 36 | 62 | 62 | 63 |
| Student performance in ELA [elem., middle, K–8 schools only] | 37 | 37 | 37 | 62 | 62 | 61 |
| Student performance in Math [elem., middle, K–8 schools only] | 40 | 42 | 39 | 58 | 57 | 58 |
| Student performance in other subjects [elem., middle, K–8 schools only] | 47 | 45 | 48 | 51 | 53 | 49 |
| Academic expectations | 37 | 41 | 35 | 62 | 59 | 64 |
| Engagement of students, parents, and educators to promote student learning | 41 | 42 | 40 | 57 | 57 | 57 |

**Table 6.1—Continued**

| Focus of School Improvement | Did Not Affect Our Focus on This Area | | | Has Caused Us to Focus Significantly or Slightly More on This Area | | |
|---|---|---|---|---|---|---|
| | All Members | Admin | UFT | All Members | Admin | UFT |
| Communication of educational goals and opportunities for feedback | 41 | 41 | 40 | 58 | 59 | 58 |
| Student performance on Regents Exams [high schools only] | 41 | 47 | 36 | 59 | 53 | 63 |
| Credits earned by students [high schools only] | 39 | 44 | 35 | 61 | 56 | 65 |
| Graduation rates [high schools only] | 43 | 47 | 40 | 57 | 53 | 60 |

NOTES: Total numbers and percentage calculations for each item exclude those who responded "don't know." Response options also included "caused us to focus significantly less on this area" and "caused us to focus slightly less on this area," but only 0–1 percent of respondents in all categories selected these options.

About one-half of CC members also reported that program participation caused them to focus slightly or significantly more on the academic performance of high-achieving students and student attendance. Interestingly, they also reported that participation increased their focus on student performance in subjects other than ELA and mathematics, which are not factored into Progress Reports and which past research has found to be less of a focus in schools responding to high-stakes accountability policies (i.e., teachers tend to focus less on and decrease instruction time in these subject areas). Respondents were least likely to report this effect on school safety: 63 percent of respondents said SPBP participation did not affect their school's focus on this area.[1] As Table 6.1 illustrates, for each area, however, more than one-third of CC members reported no effect.[2]

The survey data indicate slight increases over time in the percentage of CC members reporting that SPBP had no effect on school improvement efforts. For almost all areas of school improvement that contributed to the calculation of school Progress Reports, a higher percentage of CC members in 2009–2010 reported that SPBP participation did not affect their focus on those areas than did so in 2008–2009. Statistical tests conducted on each item showed that the differences between years were statistically significant for most of the areas (the main exception appears to be areas asked only of high schools). This suggests the observed increase in the percentage of CC members reporting no effect was not likely to have been purely due to chance.[3]

---

[1] For all these items, respondents were fairly evenly split between reporting that participation caused them to focus slightly and significantly more on school improvement. Virtually no respondents reported that participation caused them to focus less on these areas. We also found no statistically significant differences between responses of administrators and UFT CC members, suggesting general consensus on this topic.

[2] Twenty-five percent of respondents chose "did not affect" on all items listed in this question—an increase from the 21 percent in 2009.

[3] See Appendix A, p. 22, for the cross-year comparisons.

### CC Members Were More Divided About the Extent to Which Participating in SPBP Led Them to Devote More Attention and Resources to Specific School Improvement Strategies

More than one-half of all CC members reported that SPBP participation neither increased nor decreased the amount of attention and resources their schools invested in the range of improvement strategies listed in Table 6.2, such as increasing instructional time for all students, increasing opportunities to meet and work together, increasing the quantity of teacher professional development, providing before- or after-school or weekend programs, or teaching test-taking strategies to students. It is important to note, however, that schools may not have had the discretion to implement some of these strategies even if they had had the desire to do so. A very slight majority (50 to 51 percent) reported that program participation caused them to invest slightly to significantly more in several strategies, including improving response rates on Learning Environment surveys,[4] increasing the use of student achievement data to inform instruction; providing additional instruction to low-achieving students; and in the case of high schools, implementing a credit recovery program in the 2009–2010 school year.[5] Thirty-five percent of respondents chose "did not affect" on all items listed in this question.

The overall percentage of CC members reporting that SPBP participation did not affect their investment in a designated school improvement strategy increased from 2008–2009 to 2009–2010 on all 11 school improvement strategies. Statistical testing showed that the cross-year differences were statistically significant for 5 out of 11

[4] Data indicate that all schools throughout the district may have been working to improve participation in these surveys. Response rates on the annual New York City school survey increased districtwide from 55 percent in 2007–2008 to 59 percent in 2008–2009; gains were even greater for teachers, rising from 61 percent to 73 percent during this one year period (NYCDOE, 2009). We were unable to find comparable data for 2009–2010.

[5] Once again, very few respondents reported that participation caused them to invest less in these strategies. Also, the group we categorize as reporting that the effect of investing was "slightly or significantly more" was fairly evenly split between those reporting slightly more and those reporting significantly more. We also found no statistically significant differences between responses of administrators and UFT members, suggesting general consensus on this topic.

**Table 6.2**
**Percentage of CC Members Reporting Participation in SPBP Affecting Investments in School Improvement Strategies (2009–2010)**

| | Not Employed | | | Did Not Affect Our Investment in This Strategy[a] | | | Has Caused Us to Invest Significantly or Slightly More in This Strategy[a] | | |
|---|---|---|---|---|---|---|---|---|---|
| **School Improvement: Strategy** | **All** | **Admin** | **UFT** | **All** | **Admin** | **UFT** | **All** | **Admin** | **UFT** |
| Increasing instructional time for all students (e.g., by lengthening school day/year, shortening recess) | 13 | 10 | 15 | 66 | 63 | 69 | 31 | 35 | 28 |
| Implementing a credit–recovery program(s) [High schools only] | 4 | 0 | 6 | 49 | 56 | 44 | 50 | 41 | 56 |
| Improving response rates on Learning Environment survey | 4 | 2 | 5 | 47 | 51 | 45 | 51 | 48 | 53 |
| Increasing opportunities/time for teachers to meet and work together | 4 | 2 | 5 | 56 | 56 | 57 | 42 | 43 | 41 |
| Increasing the quantity of teacher professional development | 4 | 2 | 6 | 57 | 57 | 58 | 40 | 41 | 40 |
| Teaching test–taking strategies to students | 2 | 2 | 2 | 52 | 55 | 50 | 47 | 43 | 50 |
| Providing before– or after–school, or weekend programs | 3 | 3 | 3 | 56 | 59 | 54 | 42 | 39 | 44 |
| Increasing the use of student achievement data (e.g., city periodic assessments) to inform instruction | 2 | 2 | 3 | 48 | 52 | 46 | 51 | 46 | 54 |

**Table 6.2—Continued**

| School Improvement: Strategy | Not Employed | | | Did Not Affect Our Investment in This Strategy[a] | | | Has Caused Us to Invest Significantly or Slightly More in This Strategy[a] | | |
|---|---|---|---|---|---|---|---|---|---|
| | All | Admin | UFT | All | Admin | UFT | All | Admin | UFT |
| Matching curriculum and instruction with standards and/or assessments | 2 | 2 | 2 | 53 | 55 | 51 | 46 | 44 | 46 |
| Organizing Inquiry Teams of teachers and administrators to identify and address the needs of struggling students | 3 | 2 | 4 | 48 | 53 | 46 | 49 | 45 | 52 |
| Providing additional instruction to low–achieving students | 3 | 2 | 4 | 53 | 54 | 52 | 46 | 44 | 47 |

NOTES: Total numbers and percentage calculations for each item exclude those who responded "don't know."

[a] Percentages in these columns were calculated based on those who reported they had employed any of those listed school improvement strategies. Response options also included "caused us to focus significantly less on this area" and "caused us to focus slightly less on this area," but only 0–2 percent of respondents in all categories selected these options.

school improvement strategies, including increasing instructional time for all students, improving response rates on the Learning Environment survey, increasing the quantity of teacher professional development, increasing the use of student achievement data, and providing additional instruction to low-achieving students. One potential explanation is that, because so many schools received the bonus in 2008–2009, the motivational effect of the bonus decreased.[6]

Further, the vast majority of respondents reported using each of the school improvement strategies listed in Table 6.2 (a finding similar to other recent research that finds schools often respond to accountability pressures by trying multiple improvement strategies; see, for example, Padilla et al., 2006; Stecher et al., 2008). In fact, 76 percent of respondents reported that their schools used all these strategies.

Finally, although not included in Table 6.2 (because the question was worded slightly differently than recorded in the table), approximately 40 percent of CC members reported that the quality of professional development offered in the school changed slightly or significantly for the better as a result of the school's participation in SPBP. Administrators were significantly more likely than UFT members to report this positive effect, by 46 percent to 36 percent.

### Many Individuals in Case-Study Schools Reported That the Program Did Not Affect School Activities, with Two Exceptions Involving "Little Stuff"

In contrast to many survey respondents, staff across case-study schools commonly reported that they would have undertaken various changes in practices or new improvement efforts regardless of their SPBP participation. One assistant principal commented, "I don't think [the program] has affected us. It's what we have always been doing. It's not motivated by money and getting the bonus. It's motivated by doing the right things for students." Similarly, a teacher explained, "I see us doing things specifically to increase our performance, period, irrespective of a bonus." Only a handful of individuals across both years of visits acknowledged that the bonus had an effect on school activities,

[6] See Appendix A, p. 23, for the cross-year comparisons.

and they most often characterized the magnitude of this effect as quite small. A few believed the bonus helped reinforce efforts already under way. A very small minority identified specific activities undertaken in part to attain the bonus.

One possible explanation for the difference between case-study and survey respondents is the way questions were asked. During case-study interviews, respondents were asked to first describe their major school improvement efforts and then, as a follow-up, whether SPBP participation affected these efforts. In contrast, survey questions asked respondents to indicate whether participation in the program affected their school improvement efforts in a list of areas. It is also possible that survey respondents were more inclined to report socially desirable responses, or what they thought evaluators wanted to hear about the program, than were those interviewed in person.

There were, however, two common exceptions to case-study patterns of no reported effects on school activities; one individual characterized these as "the little stuff." Consistent with survey results reported in Table 6.2, interviewees commonly reported undertaking various schoolwide efforts surrounding the school environment surveys. Some of these efforts were clearly desirable and fit in with what policymakers had in mind when designing SPBP and the broader accountability system. For example, one school created a staff committee to respond to the communication problems identified in survey results, and this committee developed new structures to improve communication between and among staff (e.g., grade-level leadership positions). In contrast, some schools responded to the school environment surveys in less desirable ways that clearly worked against the intent of the accountability system and SPBP. For example, in one school, staff members reported that the administration had explained to them that the school had not received the bonus in year 1 because of the negative survey results and had encouraged them to provide more positive responses on the survey in year 2. In many schools, we also heard about significant investments in improving response rates to these surveys (e.g., educating parents on how to fill out surveys, providing multiple reminders to staff and parents, offering incentives to parents and sec-

ondary school students), and many individuals attributed these efforts to the SPBP program and a desire to win the bonus.

Second, many schools had developed strategies to improve student attendance. In some schools, these activities were immediately visible when entering the campus (e.g., posters on walls recognizing perfect attendance). Other schools reported investing in communication efforts (e.g., calls and visits to homes) and incentives (e.g., rewards and recognition for high attendance rates). For example, in addition to the many communication and recognition efforts, one school worked hard to schedule celebrations and fun events on Fridays before long weekends to ensure that students attended school those days. Some individuals reported initiating these efforts prior to SPBP, but others clearly attributed them to SPBP.

Although schools were responding to SPBP with new efforts centering on school environment surveys and attendance, only a handful of individuals reported responding by undertaking new activities or practices specifically tied to the core of teaching and learning. For example, several respondents in one elementary school attributed to SPBP increased organization of inquiry team and other collaborative meetings to identify and address individual student problems. "I guess because of the money now, it's like a little more attention is being focused on these kids, these lower range students," acknowledged one teacher in this school, who was quick to add that the bonus was not the sole motivation for teachers. "As a school, I don't think the general thought is 'we're working hard for the bonus, we're doing more for the bonus.' I don't think that's the general mindset of the teachers." One high school principal said that SPBP helped them focus improvement efforts on one particular grade level that demonstrated the lowest level in credit accumulation on the Progress Report. This same principal reported using the Progress Report Modeler to help identify this particular need. Two other principals also found the Modeler to be useful for similar purposes; however, the remaining principals across both years did not know about this tool, knew about it but did not use it, or used it and did not find it to be helpful.

Finally, a few individuals noted that, without other school-level efforts, the bonus alone could not achieve desired outcomes. In a hand-

ful of schools, staff reported that a school would not realize the benefits of the bonus if they did not take the time to examine their Progress Report target, strategize about how to address weak areas, and continually remind staff of their goals. For some, this could not occur without added time for staff to meet and without strong leadership (we will discuss this further later).

## Effects on Individuals

Overall, teachers' and CC members' views were fairly consistent with one another on the perceived effects of SPBP on themselves, other colleagues, and students.

### According to the Majority of Teachers and CC Members, SPBP Participation Did Not Affect Them Personally

On a positive note, about one-half of teachers and two-thirds of committee members reported that participation did not increase their levels of job stress (Table 6.3). Among CC members, only 11 percent said the program changed job stress for the worse and 24 percent for the better. However, teachers were less sanguine: 23 percent reported that SPBP changed the level of job stress for the worse.

As for other effects, almost two-thirds of CC members and more than one-half of teachers reported that participation did not change their own skills and abilities. Similarly, about two-thirds of teachers and CC members also said participation did not change their motivation to perform their jobs.[7] This is consistent with another survey item, in which 92 percent of teachers and 90 percent of CC members agreed or strongly agreed that the program was a nice way to recognize staff for their hard work but did not influence the way they performed their jobs (Table 6.4).

---

[7] Once again, very few respondents reported that aspects changed slightly or significantly for the worse. Also, the group we categorize as reporting that aspects changed "slightly or significantly for the better" was fairly evenly split between those reporting slightly better and those reporting significantly better.

**Table 6.3**
**Percentage of Teachers and CC Members Reporting Individual Changes Resulting from School Participation in SPBP (2009–2010)**

| | | Changed Significantly or Slightly for the Worse | Did not change | Changed Slightly or Significantly for the Better |
|---|---|---|---|---|
| My level of job stress | Teachers | 23 | 55 | 22 |
| | All CC | 11 | 65 | 24 |
| | CC admin | 11 | 63 | 26 |
| | CC UFT | 11 | 66 | 23 |
| My own skills and abilities | Teachers | 1 | 59 | 40 |
| | All CC | 0 | 65 | 34 |
| | CC admin | 0 | 64 | 36 |
| | CC UFT | 0 | 66 | 33 |
| My motivation to perform my job well | Teachers | 2 | 61 | 37 |
| | All CC | 0 | 67 | 32 |
| | CC admin | 0 | 64 | 36 |
| | CC UFT | 1 | 69 | 30 |

NOTES: Statistically significant differences between CC administrators and UFT staff are indicated in bold ($p < 0.05$). Total numbers and percentage calculations for each item exclude those who responded "don't know."

Staff members interviewed at case-study schools frequently echoed these sentiments. For example, one teacher at a school that did not win the bonus commented,

> I'm totally indifferent. It's nice but it's not what drives me on a day-to-day basis. It's nice to have a goal and see how you compare to your peers, in terms of scores. But at the end of the day, with or without the bonus plan, you still have to do your job.

As noted earlier, many individuals cited other, more-important sources of motivation. "The money had no impact on my work ethic or anything I do," said one high school teacher whose school had won the bonus the previous year. "My kids working in my class, that makes me happy. My kids being productive and excelling and learning and becoming good citizens, that's what's important to me. If I were a

**Table 6.4**
**Percentage of Teachers and CC Members from Schools That Earned Bonuses Agreeing or Strongly Agreeing with Statements About Effects of Receiving the Bonus (2009–2010)**

| | | Compensation Committee | | |
|---|---|---|---|---|
| | Teachers | All Members | Admin | UFT |
| Receiving the bonus was a nice acknowledgement of my effort but did not influence the way I perform my job this year | 92 | 92 | 92 | 92 |
| After receiving the bonus I worked harder this school year | N/A | 25 | 28 | 24 |

NOTES: Statistically significant differences between CC administrators and UFT staff are indicated in boldface ($p < 0.05$). Total numbers and percentage calculations for each item exclude those who responded "don't know." See Appendixes A and B for more detailed results on these items.

money person, I don't think I would have gone into teaching to begin with." Consistent with past research (Lortie, 1975), many individuals in this group explained that, while the bonus likely motivated others in the school, it did not affect them in this way.

A minority of individuals interviewed, however, acknowledged slight or subtle motivational effects. One principal said that the program "challenged him to push for better results." A teacher in the same school admitted, "When I am feeling down, I say, 'The bonus is there, the bonus is there.' It's not day to day, but when I am drained and tired . . . I say if you don't do what you are supposed to do, there will not be a bonus." Similarly, a teacher from another school explained, "In the back of my mind, I feel that I should work a bit harder. The more you work, the more chance there is of getting the [bonus]." Yet another teacher said, "Maybe I'm motivated [by the thought of] 'you know what, if it's available, we may as well try and get it.' I don't know. That may be part of it."

Although individuals in most schools did not report behavioral changes resulting from the program, several individuals in one school who were eligible for a share of the "incentive pool" by meeting a set of activity or behavioral criteria did report some effects. This school's

principal felt strongly that some staff, particularly nonteachers, were motivated to participate in the list of activities and had not otherwise contributed in these ways. One teacher observed many more individuals starting to write grants and initiating programs. And two teachers acknowledged not taking extra time off or missing days of work so that they would achieve the 90-percent attendance that, in part, would qualify them for a share of the incentive pool.

Cross-year analyses did not find any significant changes in the percentages of CC members reporting individual changes resulting from SPBP.[8] (See Appendix A for the cross-year comparisons.)

In schools that did not receive a bonus in year 2, the majority of teachers and CC members attribute neither positive nor negative effects to not receiving the bonus (Table 6.5). Only 8 percent of teachers and 4 percent of all CC members stated that not receiving a bonus had reduced their motivation to work harder during the year. However, unlike teachers, a significant minority of CC members reported a positive motivational effect: More than one-third reported that not receiving a bonus energized them to improve their practices in the 2009–2010 school year.

**Table 6.5**
**Percentage of Teachers and CC Members Who Did Not Earn Bonuses Agreeing or Strongly Agreeing with Statements About Effects of not Receiving the Bonus (2009–2010)**

| | | Compensation Committee | | |
|---|---|---|---|---|
| | Teachers | All Members | Admin | UFT |
| Not receiving a bonus energized me to improve my practice this year | 15 | 39 | 38 | 40 |
| Not receiving a bonus reduced my motivation to work harder this year | 8 | 4 | 3 | 5 |

NOTES: Statistically significant differences between CC administrators and UFT staff are indicated in boldface ($p < 0.05$). Total numbers and percentage calculations for each item exclude those who responded "don't know." See Appendixes A and B for more detailed results on these items.

[8] See Appendix A, p. 25, for the cross-year comparisons.

In case-study schools that did not win the bonus, teachers commonly told us that they were unaffected by this outcome. One special education teacher explained, "Not getting the bonus, I don't think it changed my philosophy. I just know that my job is to teach the kids. If I got the bonus great; if I didn't, it will not change my strategy."

### Teachers and Committee Members Generally Reported That SPBP Did Not Affect Staff Recruitment and Retention

About two-thirds or more of teachers and CC members reported that their own motivation to stay at their schools did not change as a result of program participation (Table 6.6). CC members were just as likely to report that the school's ability to recruit (70 percent) and retain (70 percent) staff did not change as a result of participating in SPBP (these items were not included in the teacher survey).

**Table 6.6**
**Percentage of Teachers and CC Members Reporting Changes in Recruitment and Retention Resulting from School Participation in SPBP (2009–2010)**

| | | Changed significantly or slightly for the worse | Did not change | Changed slightly or significantly for the better |
|---|---|---|---|---|
| My motivation to stay at this school | Teachers | 5 | 65 | 30 |
| | All CC | 2 | 70 | 28 |
| | CC admin | 1 | 69 | 30 |
| | CC UFT | 3 | 71 | 26 |
| Our ability to recruit new staff for our school | Teachers | N/A | N/A | N/A |
| | All CC | 2 | 70 | 28 |
| | CC admin | 2 | 65 | 33 |
| | CC UFT | 2 | 73 | 25 |
| Our ability to retain staff at our school | Teachers | N/A | N/A | N/A |
| | All CC | 2 | 70 | 28 |
| | CC admin | 2 | 65 | 33 |
| | CC UFT | 2 | 73 | 25 |

NOTES: Statistically significant differences between CC administrators and UFT staff are indicated in boldface ($p < 0.05$). Total numbers and percentage calculations for each item exclude those who responded "don't know."

In case-study visits, very few individuals attributed the desire to remain at their current school or to seek out another school either to the SPBP program generally or to the fact they had won or lost the bonus. "The bonus will never determine [my decision to stay]. . . . It's not big enough to impress anyone," said one teacher. Many were quick to point out other factors that were likely to attract and retain teachers to a school, including a collaborative community, positive school culture, school leaders, the community, having an A grade, parking, and offering procession pay (compensation for participating in after-school activities).

Three principals, however, reported that SPBP had positively affected either their ability to recruit new staff or to send messages about performance to their staff members. For example, one principal of a school that decided to differentiate based on performance hoped that staff receiving a smaller bonus share would reflect on their "fit" with the school and its high expectations. Although this principal could not attribute several staff member decisions to leave the school in 2008–2009 solely to SPBP, she noted that their receipt of a smaller bonus share may have contributed.

Cross-year analysis did not find significant differences in the percentages of CC members reporting changes in their ability to recruit or retain staff at their schools.[9]

### Although Teachers and CC Members Questioned the Effect on Staff Relationships, Many Saw Positive Changes in Teachers' Willingness to Collaborate, Teachers' Focus on Student Learning, and Staff Morale

Although NYCDOE and UFT leaders believed the program created more collaboration in participating schools, more than one-half of all teachers and CC members reported that participation did not change teachers' relationships with administrators, other nonclassroom staff, or other teachers (Table 6.7).

Results from case-study schools echo these findings. Aside from a few individuals, most interviewees did not report that SPBP positively

[9] See Appendix A, p. 26, for the cross-year comparisons.

affected staff interactions. In general, most individuals either reported no effect at all (e.g., "we already work together a lot") or slight negative

**Table 6.7**
**Percentage of Teachers and CC Members Reporting Changes to Staff and Staff Collaboration Resulting from School Participation in SPBP (2009–2010)**

| | | Changed significantly or slightly for the worse | Did not change | Changed slightly or significantly for the better |
|---|---|---|---|---|
| Teachers' relationships with administrators | Teachers | 9 | 57 | 34 |
| | All CC | 3 | 61 | 37 |
| | CC admin | 1 | 58 | 42 |
| | CC UFT | 4 | 63 | 33 |
| Teachers' relationships with other teachers | Teachers | 4 | 58 | 38 |
| | All CC | 2 | 57 | 41 |
| | CC admin | 1 | 54 | 45 |
| | CC UFT | 2 | 59 | 39 |
| Teachers' relationships with other non–classroom staff (e.g., counselors) | Teachers | 4 | 64 | 32 |
| | All CC | 1 | 62 | 38 |
| | CC admin | 1 | 58 | 42 |
| | CC UFT | 1 | 64 | 35 |
| Teachers' willingness to collaborate and work together | Teachers | 2 | 55 | 43 |
| | All CC | 1 | 51 | 48 |
| | CC admin | **1** | **47** | **53** |
| | CC UFT | **1** | **54** | **45** |
| Morale of school staff | Teachers | 14 | 44 | 42 |
| | All CC | 6 | 51 | 43 |
| | CC admin | **3** | **48** | **49** |
| | CC UFT | **7** | **53** | **39** |
| Teachers' focus on student learning | Teachers | 1 | 54 | 45 |
| | All CC | 0 | 52 | 48 |
| | CC admin | 1 | 49 | 51 |
| | CC UFT | 0 | 53 | 46 |

NOTES: Statistically significant differences between CC administrators and UFT staff are indicated in boldface ($p < 0.05$). Total numbers and percentage calculations for each item exclude those who responded "don't know."

effects, particularly immediately following the school vote on participation or the distribution of the bonus (e.g., in one school that differentiated bonus shares, several staff members stopped talking to one of the CC members, and the principal admitted that the experience put "a crack" in administrator-staff relations).

Nevertheless, on a positive note, almost one-half of teachers (43 percent) and CC members (48 percent) reported that teachers' willingness to collaborate and work together changed slightly or significantly for the better as a result of SPBP participation. Similarly, 42 percent of teachers and 43 percent of CC member survey respondents reported that staff morale changed slightly or significantly for the better as a result of program participation. Interestingly, administrators were significantly more likely than UFT CC members to report positive effects on staff morale (49 percent of administrators reported that staff morale changed for the better, compared with 39 percent of UFT committee members) and on teachers' willingness to collaborate (53 percent compared with 45 percent). Further, almost one-half of teachers (45 percent) and CC members (48 percent) reported that teachers' focus on student learning had changed slightly or significantly for the better.

According to one teacher in a case-study school that won the bonus, participating in the program (and earning the bonus), "makes the work more purposeful because you understand that at the end of the day, you're going to be rewarded for all of your hard work. Whereas before . . . no one recognize[d] all that hard work." In another school that won a partial bonus in year 1 and a full bonus in year 2, staff believed the program created a sense of collective pride and boost in confidence. According to the principal, "Once you get that feeling of being good at something, it doesn't go away. So I think with the teachers too [they say to themselves], 'Wow, we're good enough to get this? We're good enough!' It's a cycle."

In contrast, in some of the case-study schools that differentiated bonus distribution, staff members reported negative effects on staff morale and relations. One paraprofessional who received a smaller share than teachers reported that the bonus created conflict among staff members and that she was upset that committee members "rated"

her and gave her less without explicit criteria to defend the decision. The principal of this school acknowledged that this experience "damaged" staff-administrator relations a little, created some "trust issues," and more clearly defined "the line between us and them." In another school, several individuals reported that morale of school aides declined because they had not received a share of the bonus.

There were no statistically significant differences from 2009 to 2010 in the percentages of CC members reporting changes to staff and staff relationships resulting from SPBP.[10]

### About One-Half of Teachers and CC Members Reported Positive Effects on Students

Teachers and committee members were fairly evenly split between those who felt that student performance had not changed as a result of program participation and those who felt that it had changed slightly or significantly for the better (Table 6.8). Overall, a greater percentage of CC members reported no change in student performance in 2009–2010 than that in 2008–2009. The increase in the percentage of CC members reporting no change is greater for CC administrators than that for CC UFT members. However, this cross-year analysis showed that the differences between years and roles are close yet not statistically significant.

### Teachers in Schools That Won the Bonus Were More Likely Than Those in Schools That Did Not to Report Positive Changes in Staff Morale, Their Own Motivation to Stay at Their Schools, and Improvements in Student Performance

A few perceptions about the effects of SPBP differed among teachers in schools that won the bonus in year 2 and those that did not. A significantly higher percentage of teachers in the bonus schools than in the nonbonus schools reported that, as a result of SPBP, the following had changed slightly or significantly for the better: morale of school staff

[10] We did, however, observe significantly greater increases in the percentage of CC members reporting "changed slightly or significantly for the better" between the two years among CC UFT members than among CC administrators. See Appendix A, p. 27, for the cross-year comparisons.

**Table 6.8**
**Percentage of Teachers and CC Members Reporting Student Performance Changes Resulting from School SPBP Participation (2009–2010)**

| | Changed significantly or slightly for the worse | Did not change | Changed slightly or significantly for the better |
|---|---|---|---|
| Teachers | 4 | 47 | 49 |
| All CC | 2 | 54 | 45 |
| CC admin | 1 | 53 | 46 |
| CC UFT | 2 | 54 | 44 |

NOTES: Statistically significant differences between CC administrators and UFT staff are indicated in boldface ($p < 0.05$). Total numbers and percentage calculations for each item exclude those who responded "don't know."

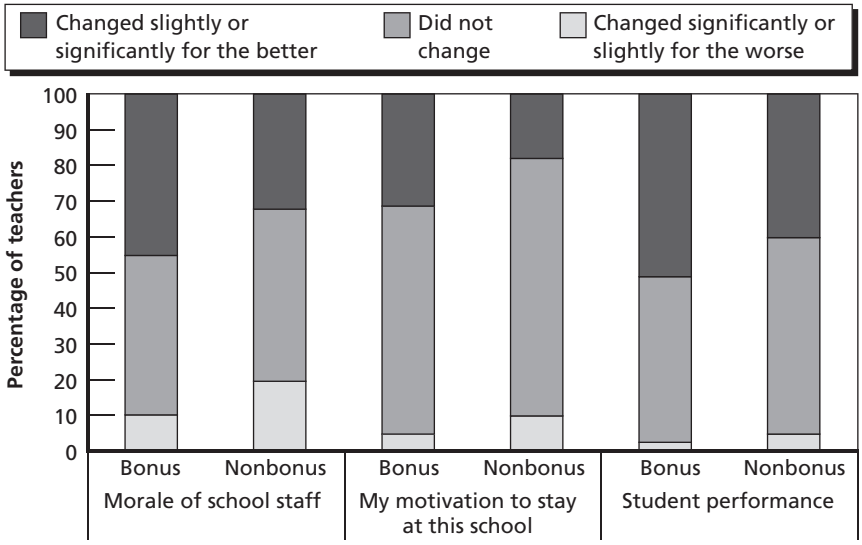
(45 percent to 32 percent), their own motivation to stay at their schools (31 percent to 18 percent), and student performance (51 percent to 40 percent). (See Figure 6.1.) Because we did not have measures from before and after these schools won the bonus, we could not determine whether winning the bonus was associated with these changes in perceptions, possibly making them a consequence of that win, or whether these differences preceded the bonus, possibly meaning that they are due to the different outcomes. It is also possible that neither explanation is valid and that teachers in the schools that did not win the bonus had more negative views in general and that these had nothing to do with the program or its outcomes.

## Summary and Discussion

District and UFT leaders, teachers, school CC members, and case-study school staff reported the bonus and program participation overall had mixed effects. Respondents also conveyed varied opinions about the effects of participating in the program.

As for perceived effects on school improvement efforts,

**Figure 6.1**
**Percentage of Teachers Reporting Changes Resulting from School SPBP Participation (2009–2010)**




RAND *MG1114-6.1*

- More than one-half of the CC members said SPBP participation led them to focus more on many areas of school improvement that contributed to the calculation of Progress Reports, including outputs (e.g., academic performance of high-needs students, student performance in ELA and mathematics) and inputs (communication of goals, engagement of stakeholders). Nevertheless, more than one-third of the CC members reported that the program did not affect their focus in these areas.
- From 2008–2009 to 2009–2010, the survey data indicate slight increases in the percentage of CC members reporting that SPBP did not affect their school improvement efforts, particularly those contributing to the calculation of school Progress Reports (e.g., focus on attendance, student performance in ELA and mathematics).
- CC member respondents were more divided in how they viewed the program's effects on specific school improvement strategies:

Approximately one-half or more said SPBP participation did not cause them to focus any more or less attention and or invest any more resources on a range of improvement strategies. Case-study schools reported SPBP did have some effect on the "little stuff"—efforts to improve school environment survey response rates and students attendance.

As for perceived effects on individuals:

- The majority of teachers, CC members, and case-study respondents said SPBP did not affect them personally, including their job performance, skills, and motivation.
- Winning and losing the bonus was not viewed as being very influential for job performance or motivation. The vast majority of teachers and CC members said that winning the bonus was a nice acknowledgement of their hard work but did not affect their performance. Less than 10 percent of teachers and CC members whose schools did not earn the bonus said that not receiving the bonus reduced their motivation. Unlike teachers, however, a substantial minority of CC members (more than one-third) reported that not receiving a bonus had energized them to improve their teaching practices the subsequent year.
- Most respondents also believed the program did not change teachers' relations with other teachers, administrators, or nonclassroom staff, and they did not typically observe effects on staff recruitment or retention.
- However, more than 40 percent of teachers and CC members reported that teachers' willingness to collaborate and work together, teachers' focus on student learning, and staff morale changed slightly or significantly for the better as a result of SPBP participation (although the numbers for CC members were driven in part by the responses of administrators, who were significantly more likely than UFT-represented committee members to report positive effects on staff morale and willingness to collaborate).
- Teachers and CC members were divided in their perceptions of how SPBP had affected students: Approximately one-half believed

student performance had not changed, and one-half believed it had changed slightly or significantly for the better.

Taken together, some of these findings appear to be contradictory—particularly those on the reported motivational effects of the bonus. For example, although the majority of respondents said that winning the bonus did not influence their job performance or motivation, more than one-half reported that SPBP participation led them to focus more on many areas of school improvement. It is possible that respondents may have been responding individually to the question of how winning the bonus affected job performance or motivation but were responding more for their school as a whole to the question of how participation affected the focus of school improvement. Thus, while they may not have felt the effects individually, on their own motivation, instruction, or practices, they may have believed that school staff or other colleagues were shifting focus in certain ways. In some respects, respondents may have been distancing themselves from claims that financial incentives influenced them personally, while also acknowledging that the incentives may be motivating for other colleagues (Lortie, 1975, found similar findings).

The seemingly contradictory results may also reflect the internal conflicts individuals experience when thinking about and responding to questions about bonuses. As we found in the case-study visits, some individuals contradicted themselves in interviews, commenting at one point that they were opposed to the notion of merit pay, yet supportive of the idea that individuals contributing more to school success deserved a greater share of the bonus pool. Similarly, some strongly rejected the assertion that financial incentives would motivate them to do anything different from what they already do, yet, at the same time, they indicated that the prospect of receiving a bonus was a useful incentive. This, too, is consistent with survey results indicating that teachers did not believe receiving a bonus changed their practices but, as discussed in Chapter Four, considered it a potential motivator, suggesting that it was something they kept in the back of their mind as they came to work each day. They did, therefore, appear to be taking the bonus seriously.

Finally, in the aggregate, these reported effects are a useful indicator of how participating staff members viewed the program. Although the perceptions of the effects may not correlate with direct estimates of the effects (as reported in the next two chapters), they provide a valuable measure of the underlying level of support for SPBP. The findings in part help answer the question of whether staff members were open enough to the program for it to succeed or whether attitudes were so negative that the program held no possible chance of success. Staff willingness to attribute positive effects to SPBP—particularly regarding school improvement efforts—suggests that, in fact, large segments of the school populations bought into the program. However, one area in which they perceived fewer effects was their own individual behavior. This finding suggests two possible conclusions. First, it suggests that it may be very challenging to motivate individuals with financial rewards either because they resist acknowledging that they were motivated (as indicated by the earlier discussion about contradictions in how individuals respond to questions about motivation) or because they truly are not motivated by such rewards (as findings in Chapter Eight appear to indicate). Second, these findings might explain why we did not in fact see SPBP effects: Staff reported that the bonus did not personally motivate them to change their behaviors.

CHAPTER SEVEN

# Effects on Progress Report and Student Test Scores

The primary goal of SPBP was to change the education productivity by rewarding teachers and other UFT members for better student outcomes, as measured by the Progress Reports. This chapter first reviews how SPBP schools did on the performance metrics used to award bonuses: the Progress Report overall and component scores. It also examines how the program affected these metrics. The chapter then turns to more-direct measures of student achievement, investigating how SPBP affected student test scores, as measured by the State of New York's standardized assessments during the three years of implementation. The purpose of this analysis was to determine whether students enrolled in a participating school had higher achievement scores than they would have had, had the school not participated. Our approach was to compare test scores of students enrolled in schools invited to participate in SPBP with test scores for a comparable set of students and schools not invited to participate in the program.

This chapter addresses the following research questions:

1. How did SPBP affect Progress Report scores across all years and all component scores?
2. How did SPBP affect student test scores across years?
    a. What were the effects on mathematics or ELA test scores for elementary, middle, and K–8 students?
    b. What were the effects on Regents Exam scores for high school students?

c. What was the relationship between achievement levels of elementary, middle, and K–8 school students and their schools' implementation of the bonus program, as measured by the differentiation of the bonus awards?

The lottery selection process that Roland Fryer implemented for NYCDOE to identify the schools to invite to participate in SPBP helped ensure that the characteristics of the participating schools would be comparable to those of the schools entered into the lottery but not invited to participate. Since schools that both met the definition of "high needs" and were entered into the lottery had an equal chance of being invited to participate in SPBP, we could attribute differences in student outcomes to SPBP participation rather than to some other, preexisting difference between the two groups of schools.

In the remainder of this chapter, we first describe the data used for our analyses. We then discuss Progress Report scores and individual student test score analyses and findings.

## Data

The analysis of how SPBP influenced Progress Report scores and student test results required school-level and student-level data, including student characteristics and student test scores. The Progress Report analyses relied on the school-level data, while the student test score analyses relied on both sets of data to identify and then analyze test results for the students enrolled during the year just prior to SPBP and all three years of the program in schools that were entered into SPBP lottery in fall 2007.[1] We discuss each type of data in turn.

### School-Level Aggregate Data and Progress Report Data

NYCDOE provided a series of data files on school characteristics for the four school years of interest. The files contained demographic infor-

[1] Thus, our data derive from the 2006–2007, 2007–2008, 2008–2009, and 2009–2010 school years.

mation on the students enrolled in each school, including attendance rates, total enrollment, and disciplinary incidents. The files also contained AYP status and ratings assigned to the school, as defined by the State of New York's accountability program under NCLB. As noted in Chapter Two, for the analysis of the Progress Reports, we obtained the most current Progress Report scores and grades and the component scores and grades from the NYCDOE website.

### Student-Level Data

In addition to the school-level data, the analysis of student test scores required additional data on student characteristics and student test scores.

**Student Biographical Data.** NYCDOE provided two types of student-level data files: student biographical information and student achievement test scores. The student biographical files contained information on the universe of students enrolled in New York City public schools by school year and the school and grade in which each student was registered by semester. The biographical files also contained student background information, including gender, race and/or ethnicity, participation in special education services and ELL programs, and eligibility for free or reduced-price school lunches. Both the biographical and achievement data files include a unique, longitudinally consistent identifier for each student. This means that records on the same student can be linked across multiple schools and time periods and to records contained in separate data files.

**Student Test Score Data.** NYCDOE provided a series of data files containing test scores for the universe of students enrolled in elementary, middle, and K–8 schools. The department also provided a similar set of files for high school students, although these files lacked standardized variable names and definitions from one school year to the next. For grades 3 through 8, the student test files contained scores from the State of New York's mathematics and ELA assessments administered during the year just prior to SPBP and for all three years of the program. The high school student test files contained scores from the Regents Exams, Regents Competency Tests, Regents Portfolio Assessment, and approved alternative examinations for the year just prior to

SPBP and for the first two years of the program. High school data for year 3 (the 2009–2010 school year) were not available at the time we prepared this report.

We standardized the state test scale scores of students enrolled in elementary, middle, and K–8 schools by subject, grade, and school year to ease threats associated with potential variations in content standards, grade-level expectations, test constructs, performance standards, etc. Using data for the universe of students enrolled in the NYCDOE school system, we estimated the means and standard deviations of scores by grade level for each school year and subject. The standardized score equals the quantity of the scale score minus the mean for the corresponding grade level, subject, and year divided by the standard deviation for the grade level, subject, and year. If we let $z$ denote standardized score, $y$ denote the scale score, and $m$ and $s$ denote the mean and standard deviation of the scores of all students for the corresponding grade level and school year, then

$$z = \frac{y - m}{s}.$$

We used both the standardized z-scores and the scale scores for all analyses presented in this chapter.

Standardized z-scores have two potential advantages over the scale scores. First, estimated SPBP effects for z-scores are in terms of standard deviations and can be considered effect sizes, which can be compared with the effects of other programs, such as promotion policies.[2] Second, although the scale scores are on a common scale across grades so that scale points are meant to have the same interpretation at every grade level, the scale might not meet the requirements of a

[2] Effect sizes are commonly used to evaluate the size of a program effect in units that can be compared across studies. However, effects that are equal in standard deviation units are not necessarily equal in terms of other quantities of interest. For instance, the scale score points between proficiency levels from two tests may not be equal in standard deviation units. Effects that are equal in standard deviation units would not be equal in shifting the proportion of students changing proficiency level. Moreover, effect sizes are sensitive to the sample used to calculate the standard deviation. For these reasons, we provide estimates in both z-score and raw score units.

developmental (or interval) scale (compare Allen and Yen, 1979) across grades, and points from one grade might have different meaning than points in another grade level. If this is true, an estimated SPBP effect that combined scale score data across grades would be difficult to interpret. The z-scores from each grade are in terms of the standard deviation of the grade, and these metrics are comparable across grades, even if the scale score points are not.

However, given the distribution of scores in each grade and where those scores are relative to important standards, such as proficiency, it could be argued that a difference of one standard deviation has different meanings across grades or that the effort required to change scores in standard deviation units will differ for students at different grade levels. SPBP effects would then not be constant in standard deviation units across grades. If this is true, our estimated effect is a weighted average of the grade-specific effects, where weights are relative numbers of students in the sample from each grade level.[3] We cannot test whether the scale score points are truly comparable across grades or whether z-scores are. Hence, we cannot be sure that estimated effects that pool data across grades are not sensitive to the scaling of the scores. We therefore have provided estimates using both scale scores and z-scores and have conducted additional sensitivity analyses on the test scales

We did not report estimated effects for ELA tests for elementary or middle grade students in year 1 because these assessments were administered around the same time schools were deciding whether to participate in SPBP, leaving little or no time for participation in the program to affect student achievement.[4]

For students enrolled in grades 9 through 12, the State of New York uses the Regents Exams to assess their academic performance. Regents Exams focus on individual subject areas, including English,

---

[3] The weighting of different grade levels between the SPBP and control samples will be dependent on the schools assigned to each group by the lottery and could change with different potential assignment contributing to additional variance in our estimated effects.

[4] We did estimate and test this effect, and it was small and not statistically significant. ELA tests were administered in January, and schools were invited to participate in September, but some schools did not report voting and notifying the district of their decision to participate until December and January.

foreign languages, mathematics, science, and social studies. All Regents Exams are offered in June of each school year, with a limited number of examinations being offered in January and August. Students can choose when to take a Regents Exam and can retake the test if they did not pass at the level required for graduation.

To receive a Local High School Diploma, students in New York City starting high school in the 2005–2006 or 2006–2007 school year must pass five exams with a score of 55 or higher: Mathematics (including Mathematics A or Integrated Algebra), Global History and Geography, U.S. History and Government, Comprehensive English, and any one science Regents Exam. To receive a Regents Diploma, students must pass with a score of 65 or higher; to receive a Regents Advanced Diploma, they must meet the requirements of the Regents Diploma plus score 65 or higher on additional exams, including an additional Regents science exam and a foreign language exam. Students starting high school in the 2008–2009 school year and beyond must meet the performance standards defined by the Regents to receive a diploma (the Local High School Diploma is available only to special education students).

We estimated separate SPBP effects for six Regents Exams: two required exams, Comprehensive English and Mathematics A (Integrated Algebra in 2009); Science Living Environment, the exam most commonly taken to fulfill the science exam requirement; and Mathematics B, Earth Science, and Chemistry, the other exams most commonly taken by students in SPBP and control schools according to the data we received from NYCDOE. We also estimated the effects separately for the fall (year 2) and spring (years 1 and 2) test administrations. We did not estimate effects for the year 1 fall administration because exams were administered before schools voted on participation in the program. We also did not estimate effects for year 3 because these test scores were unavailable before we completed our analysis. Additionally, we did not estimate SPBP effects separately by grade level for students enrolled in grades 9 through 12, since grade-level information was incomplete. Students who were retaking a particular Regents Exam were identified by exam-specific form codes denoting a retest situation, a student having multiple test scores for the same subject in a

single year, and a student having multiple test scores for the same subject across school years. We restricted the analysis sample to first-time test takers.

Similar to the analysis of student test scores for grades 3 through 8, we estimated SPBP participation effects using both the raw values and standardized z-scores. Regents Exams are scored on a 100-point scale. Exam scores ranged from 1 to 100, and districtwide the standard deviations of the exams were generally around 30.0 points. We standardized the Regents Exams scores by test subject and test administration using the mean and standard deviation from schools entered into SPBP lottery. We did not standardize Regent Exams using data for the universe of students enrolled in the school system due to incomplete information for a large number of observations.

We extensively cleaned the data files to identify anomalous enrollment patterns and correct missing or longitudinally inconsistent biographical information. We also made variable definitions and labels consistent across years to permit multiyear evaluation of SPBP. For example, because we could link student records across multiple years and files, we were able to fill in missing time-invariant information on students. We then created longitudinal records with multiple years of test score results so that student performance on these assessments from the year prior to a school participating in SPBP could be incorporated in the analysis plan.

## Basic Analytic Approach

A simple estimate of SPBP's effect would be the difference between the average test score for students enrolled in participating schools and the average test score for the students enrolled in eligible schools not participating. Ideally, this approach would allow us to infer how much average student test scores would differ if their schools did or did not participate in SPBP.

However, as discussed in Chapter Two, not all schools randomly assigned to SPBP by the lottery elected to participate. Staff in some schools voted not to participate after being invited, while other schools

dropped out during or after the first three years of implementation. Also, some schools not randomly assigned to SPBP participated in the program. Schools' ability to choose whether to participate or not and the fact that some schools joined the program without being randomly assigned present challenges to our analysis because schools that did not follow their lottery assignments might differ systematically from other schools in the sample in ways that are related to student test scores. If these subsets of schools are meaningfully different, comparing the test scores of students in schools that did and did not participate in SPBP can distort SPBP effects. That is, if we use assignments that schools chose rather than assignments controlled through the lottery, we might attribute differences among schools that do and do not choose SPBP participation to the effects of the program.

Rather than use a simple analytic approach and risk unknown errors, we implemented an alternative approach that is commonly used in estimating treatment effects in randomized control trials, the ITT analysis (Shadish, Cook, and Campbell, 2001). It compares the average outcomes, i.e., Progress Report scores or student test scores, of schools or students enrolled in schools that were assigned by lottery to SPBP, regardless of whether the school participated in the program in a particular school year, to the average outcomes for schools that were not selected by the lottery process to SPBP, regardless of whether the school participated in SPBP in a particular school year.

An ITT analysis may seem counterintuitive. Students enrolled in some schools that did not participate in SPBP are included in the treatment group, while students from some schools that did participate are included in the control group. Although this will result in errors in the estimate of the SPBP's effect on student test scores, the direction of the bias is known. The estimated effects will take on the same sign, but the magnitude of the treatment effect will be closer to zero than the true effect (attenuation bias). An ITT analysis also supports significance testing of the hypothesis that the effect of SPBP is not different from zero.

Furthermore, under the appropriate assumptions, which are reasonable for SPBP evaluation, we can apply a statistical technique developed by Bloom (1984) and others (Greene, 2007) to generate estimates

from the ITT analysis that correct for the attenuation bias. Heuristically, this adjustment is achieved by dividing the SPBP effect from the ITT analysis by the fraction of treatment schools that actually voted in favor of participation. We conducted the adjusted ITT analyses for our student-level results and obtained results that were very similar to the unadjusted results because the overall participation rate among schools invited to participate in SPBP was relatively high (over 84 percent). Consequently, we only present the unadjusted results in this chapter, the adjusted results appear in Appendix D.

Also as discussed in Chapter Two, NYCDOE dropped performance targets for schools not participating in SPBP in years 2 and 3. This technically changes the interpretation of the SPBP effect being estimated in year 1 in comparison with that of the other years. In year 1, the effect was that of the bonus program alone, since both SPBP and other schools had targets. In years 2 and 3, the estimated effect could have been due to the actual potential for financial rewards or the effect of a target for performance. However, the target was only for the financial rewards, so this distinction may be limited. Also, we expect that a minimal effect on results from the distinction of whether or not control schools had a target because, with or without explicit targets, there were high stakes associated with Progress Report scores and grades for all schools.

## Sample

This section describes the schools that were and were not invited to participate in SPBP. The first three tables show select student- and school-level sample statistics for elementary, middle and K–8 schools by year. The following four summarize similar statistics for high schools. Because of the focus on the ITT analysis, we have defined SPBP schools or treatment schools as the set of high-needs schools randomly selected to participate in the program, regardless of whether school personnel employed in the school actually voted in favor of participating in the program. Control schools are defined as the set of high-needs schools entered into SPBP lottery but not randomly selected and invited to par-

ticipate. In addition to reporting mean values for each group of schools, the tables indicate whether the differences in mean values between SPBP and control schools are statistically significant.

### SPBP and Control Sample Elementary, Middle, and K–8 Schools Are Very Similar

Our analysis sample for students in grades 3 to 8 included students from 191 schools invited to participate in SPBP and 133 control schools. The 191 SPBP schools included 114 elementary, 55 middle, and 22 K–8 schools randomly selected in the lottery. The 133 control schools included 78 elementary, 41 middle, and 14 K–8 schools that were not selected in the lottery. Due to school closings, school consolidations, and incomplete attrition, the sample of schools by year 3 decreased to 187 treatment schools and 132 control schools.

In general, the observable characteristics of the two student populations were very similar the year prior to SPBP (baseline, or year 0). As shown in Table 7.1, roughly 48 percent of the students were female, in both samples; 55 percent of the students in treatment schools and 56 percent in control schools were Hispanic. A large majority of the students also came from low-income families, with roughly 80 percent of students in treatment schools and 82 percent of students in control schools being ELL. Another 5 to 6 percent of students were eligible for reduced-price lunches in SPBP and control samples, respectively.

Table 7.1 also compares the achievement levels of students in the two groups. For each year, the table compares the average year 0 achievement of students enrolled in the SPBP schools to the average for students in the control schools. Because testing begins in grade 3 and students transfer into the district, the sample of students with year 0 scores decreased with each year of implementation from essentially all students in year 0 to only about 20 percent of students in year 3. In both the SPBP and control samples of schools, the average student scored around 652 on the mathematics test and 638 on the ELA tests in the year before the program. These scores place the average student in SPBP schools approximately one-third to one-half of a standard deviation unit below the average student in the New York City public school system as measured by the State of New York's mathematics and

**Table 7.1**
**Select Student-Level Summary Statistics for Elementary, Middle, or K–8 Schools, by Treatment Status and Year**

| | Year 0 | | | Year 1 | | | Year 2 | | | Year 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Treat. | Cont. | Diff. | Treat. | Cont. | Diff. | Treat. | Cont. | Diff. | Treat. | Cont. | Diff. |
| **Student characterisitics** (%) | | | | | | | | | | | | |
| Female | 48.42 | 48.47 | –0.04 | 48.69 | 48.11 | 0.59 | 48.90 | 48.41 | 0.49 | 48.12 | 47.82 | 0.30 |
| Male | 51.58 | 51.53 | 0.04 | 51.31 | 51.89 | –0.59 | 51.10 | 51.59 | –0.49 | 51.88 | 52.18 | –0.30 |
| Asian/Pacific Islander | 1.77 | 2.03 | –0.26 | 1.72 | 1.97 | –0.25 | 1.86 | 1.89 | –0.03 | 1.90 | 2.11 | –0.21 |
| Black | 41.03 | 39.84 | 1.20 | 40.81 | 39.85 | 0.96 | 40.08 | 39.44 | 0.65 | 39.51 | 39.60 | –0.09 |
| Hispanic | 55.46 | 56.30 | –0.85 | 55.80 | 56.40 | –0.61 | 56.29 | 56.91 | –0.62 | 56.81 | 56.25 | 0.56 |
| Native Am./ Am. Indian | 0.48 | 0.40 | 0.09 | 0.48 | 0.40 | 0.08 | 0.41 | 0.33 | 0.08 | 0.32 | 0.34 | –0.03 |
| White | 1.16 | 1.36 | –0.19 | 1.11 | 1.32 | –0.20 | 1.19 | 1.30 | –0.12 | 1.19 | 1.51 | –0.32 |
| Race undefined | 0.09 | 0.08 | 0.01 | 0.08 | 0.06 | 0.02 | 0.17 | 0.13 | 0.04 | 0.28 | 0.19 | 0.08 |
| Reduced-price lunch | 4.86 | 5.68 | –0.82 | 4.50 | 5.19 | –0.68 | 4.21 | 4.62 | –0.41 | 3.09 | 2.98 | 0.11 |
| Free lunch | 79.96 | 82.27 | –2.32 | 81.16 | 80.08 | 1.08 | 78.76 | 79.43 | –0.67 | 93.28 | 93.15 | 0.13 |
| ELL | 46.42 | 48.29 | –1.87 | 44.65 | 46.70 | –2.05 | 44.62 | 46.37 | –1.75 | 20.60 | 20.79 | –0.19 |
| Retained | 0.14 | 0.17 | –0.04 | 2.72 | 3.05 | –0.33 | 1.91 | 2.13 | –0.22 | 4.15 | 4.34 | –0.20 |
| Enrollment | 701 | 703 | –3 | 668 | 667 | 1 | 627 | 622 | 5 | 666 | 663 | 3 |
| Total students | 70,234 | 49,098 | — | 65,883 | 45,619 | — | 64,412 | 45,872 | — | 63,177 | 45,920 | — |

**Table 7.1—Continued**

| | Year 0 | | | Year 1 | | | Year 2 | | | Year 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Treat. | Cont. | Diff. | Treat. | Cont. | Diff. | Treat. | Cont. | Diff. | Treat. | Cont. | Diff. |
| Year 0 (2006–2007) Test Scores[a] | | | | | | | | | | | | |
| Scale Scores | | | | | | | | | | | | |
| Mathematics | 651.63 | 651.72 | –0.09 | 654.45 | 654.81 | –0.36 | 656.32 | 656.55 | –0.23 | 655.92 | 656.22 | –0.30 |
| ELA | 637.98 | 638.39 | –0.41 | 638.53 | 638.96 | –0.42 | 639.25 | 639.23 | 0.03 | 637.70 | 637.42 | 0.28 |
| Standardized Z–Scores | | | | | | | | | | | | |
| Mathematics | –0.37 | –0.36 | –0.01 | –0.38 | –0.37 | –0.01 | –0.41 | –0.41 | –0.01 | –0.50 | –0.49 | 0.00 |
| ELA | –0.36 | –0.34 | –0.01 | –0.37 | –0.36 | –0.01 | –0.41 | –0.40 | 0.00 | –0.47 | –0.48 | 0.01 |
| Total students | 67,540 | 47,208 | — | 45,664 | 31,534 | — | 28,135 | 20,383 | — | 13,838 | 10,581 | — |
| Total schools | 191 | 133 | 58 | 191 | 133 | 58 | 187 | 133 | 54 | 187 | 132 | 55 |

NOTE: Test scores for year 1, 2, and 3 are from the last year prior to the SPBP being implemented. Four treatment schools are missing in year 2 and four treatment schools and 1 control school are missing in year 3.

[a] For every study year the year 0 scores reported in the table are the average of the test scores from the 2006–2007 school year for students enrolled in the study schools during the study year. Because testing begins in grade 3 and students transfer into the districts, for each additional year of the program fewer of the students enrolled in study schools that year have 2006–2007 test score. This explains the decline in the sample sizes for students with year 0 scores from year 0 to year 3.

ELA assessments, which is consistent with these schools' designation as high-needs schools.

Table 7.2 displays school-level summary statistics for elementary, middle, and K–8 schools by treatment status and school year. In both samples, the average school size (about 595 students) and the distribution of schools among school types were very similar between the SPBP and control school samples. For example, approximately 60 percent of treatment schools and 58 percent of control group were classified as elementary schools. As with the summary statistics on student characteristics reported in Table 7.1, treatment and control schools appeared to be very much alike when student characteristics and proficiency levels were aggregated at the school-level. This was true not only at the baseline but also across the three years of SPBP implementation. If anything, the two groups tended to look more similar over time, which was most noticeable in the proportions of free and reduced-price lunch status students, which decreased by 3 percentage points (–3.41 to –0.29) from the year just prior to SPBP to year 3 of the program.

### The SPBP and Control High School Samples Are Similar but Have Some Differences on Background Variables

Our high school sample included students from 55 high schools invited to participate in SPBP and 42 control schools. The 55 high schools included 44 high schools, 10 middle–high schools, and one transfer high school. The 42 control schools included 35 high schools, one K–8 high school, and six middle-high schools. Due to school closings, consolidations, incomplete attrition, and missing data, the sample of schools decreased by year 2 to 47 SPBP and 33 control high schools.

The summary statistics in Table 7.3 indicate that students enrolled in SPBP schools were more likely to be black than students in schools not selected to participate. The same held true for students eligible for free school meals, although the difference in the average percentage of students qualifying for free lunch nearly disappeared by year 2. These statistics also indicate that students enrolled in SPBP schools were less likely to be Hispanic or classified as ELL. Despite SPBP and control high schools being slightly different on select background characteristics, none of the differences were statistically significant at conventional

**Table 7.2**
**Select School-Level Summary Statistics for Elementary, Middle, and K–8 Schools, by Treatment Status and Year**

| | Year 0 | | | Year 1 | | | Year 2 | | | Year 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Treat. | Cont. | Diff. | Treat. | Cont. | Diff. | Treat. | Cont. | Diff. | Treat. | Cont. | Diff. |
| **Student characteristics** (%) | | | | | | | | | | | | |
| Female | 48.57 | 48.69 | –0.13 | 48.88 | 48.18 | 0.69 | 49.01 | 48.57 | 0.45 | 48.33 | 47.85 | 0.48 |
| Male | 51.43 | 51.31 | 0.13 | 51.12 | 51.82 | –0.69 | 50.99 | 51.43 | –0.45 | 51.67 | 52.15 | –0.48 |
| Asian/Pacific Islander | 1.65 | 1.62 | 0.04 | 1.61 | 1.58 | 0.04 | 1.75 | 1.63 | 0.12 | 1.76 | 1.65 | 0.11 |
| Black | 41.15 | 43.49 | –2.34 | 41.21 | 43.21 | –2.00 | 40.66 | 42.49 | –1.83 | 40.14 | 41.92 | –1.78 |
| Hispanic | 55.41 | 53.17 | 2.25 | 55.45 | 53.60 | 1.85 | 55.69 | 54.25 | 1.43 | 56.09 | 54.55 | 1.54 |
| Native Am./ Am. Indian | 0.43 | 0.43 | 0.00 | 0.43 | 0.43 | 0.01 | 0.40 | 0.34 | 0.06 | 0.34 | 0.37 | –0.02 |
| White | 1.27 | 1.20 | 0.07 | 1.20 | 1.11 | 0.10 | 1.32 | 1.12 | 0.19 | 1.37 | 1.29 | 0.08 |
| Free/reduced lunch | 85.81 | 89.22 | –3.41 | 86.31 | 87.17 | –0.86 | 84.09 | 85.85 | –1.77 | 96.04 | 96.33 | –0.29 |
| Average enrollment (no.) | 595 | 593 | 2 | 578 | 572 | 6 | 528 | 515 | 12 | 569 | 562 | 7 |

**Table 7.2—Continued**

| | Year 0 | | | Year 1 | | | Year 2 | | | Year 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Treat. | Cont. | Diff. | Treat. | Cont. | Diff. | Treat. | Cont. | Diff. | Treat. | Cont. | Diff. |
| Proficiency | | | | | | | | | | | | |
| Mathematics (%) | | | | | | | | | | | | |
| Level 1 | 15.15 | 15.48 | –0.32 | 9.05 | 8.74 | 0.30 | 4.85 | 4.47 | 0.38 | 16.46 | 15.39 | 1.07 |
| Level 2 | 32.30 | 32.28 | 0.02 | 27.41 | 26.65 | 0.76 | 21.52 | 20.33 | 1.20 | 46.56 | 44.98 | 1.58 |
| Level 3 | 43.58 | 42.71 | 0.87 | 53.91 | 53.93 | –0.02 | 59.50 | 59.69 | –0.19 | 26.98 | 28.44 | –1.45 |
| Level 4 | 8.97 | 9.54 | –0.57 | 9.63 | 10.67 | –1.04 | 14.13 | 15.51 | –1.38 | 10.00 | 11.19 | –1.20 |
| ELA (%) | | | | | | | | | | | | |
| Level 1 | 13.31 | 13.49 | –0.19 | 8.88 | 9.15 | –0.27 | 4.18 | 4.18 | 0.00 | 23.03 | 22.36 | 0.67 |
| Level 2 | 51.76 | 50.40 | 1.36 | 48.87 | 46.83 | 2.04 | 40.30 | 38.87 | 1.44 | 50.42 | 49.85 | 0.57 |
| Level 3 | 33.83 | 34.87 | –1.04 | 40.82 | 42.56 | –1.74 | 53.45 | 54.65 | –1.20 | 23.53 | 24.60 | –1.07 |
| Level 4 | 1.11 | 1.24 | –0.14 | 1.42 | 1.45 | –0.03 | 2.07 | 2.31 | –0.24 | 3.02 | 3.19 | –0.18 |
| School Type | | | | | | | | | | | | |
| Elementary | 114 | 78 | 36 | 112 | 76 | 36 | 110 | 77 | 33 | 109 | 76 | 33 |
| Middle | 55 | 41 | 14 | 55 | 41 | 14 | 54 | 41 | 13 | 54 | 40 | 14 |
| K–8 | 22 | 14 | 8 | 24 | 16 | 8 | 23 | 15 | 8 | 24 | 16 | 8 |
| Total schools (no.) | 191 | 133 | 58 | 191 | 133 | 58 | 187 | 133 | 54 | 187 | 132 | 55 |

NOTES: Proficiency levels for years 1, 2, and 3 are from the last year prior to the SPBP being implemented. Four treatment schools are missing in year 2, and 4 treatment and 1 control school are missing in year 3.

**Table 7.3**
**Select Student-Level Summary Statistics for High Schools, by Treatment Status and Year**

| | Year 0 | | | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|---|---|---|
| **Student Characteristics** | **Treatment** | **Control** | **Difference** | **Treatment** | **Control** | **Difference** | **Treatment** | **Control** | **Difference** |
| Female (%) | 47.05 | 47.93 | –0.89 | 44.86 | 46.22 | –1.37 | 43.76 | 44.68 | –0.92 |
| Male (%) | 48.14 | 47.93 | 0.21 | 48.23 | 46.11 | 2.13 | 47.19 | 46.69 | 0.50 |
| Gender undefined (%) | 4.81 | 4.13 | 0.68 | 6.91 | 7.67 | –0.76 | 9.05 | 8.63 | 0.42 |
| Asian/Pacific Islander (%) | 4.24 | 4.44 | –0.20 | 3.96 | 4.16 | –0.19 | 3.60 | 3.49 | 0.12 |
| Black (%) | 44.55 | 35.21 | 9.34 | 44.22 | 33.42 | 10.81 | 41.60 | 33.31 | 8.28 |
| Hispanic (%) | 43.46 | 52.49 | –9.02 | 42.28 | 51.47 | –9.19 | 43.54 | 51.61 | –8.06 |
| Native Am./Am. Indian (%) | 0.40 | 0.36 | 0.05 | 0.43 | 0.33 | 0.10 | 0.46 | 0.34 | 0.12 |
| White (%) | 1.89 | 3.04 | –1.15 | 1.80 | 2.67 | –0.87 | 1.55 | 2.39 | –0.84 |
| Race undefined (%) | 5.45 | 4.47 | 0.99 | 7.30 | 7.96 | –0.66 | 9.25 | 8.86 | 0.38 |
| Reduced-price lunch (%) | 6.71 | 4.57 | 2.14 | 7.47 | 5.77 | 1.70 | 6.31 | 5.83 | 0.48 |
| Free lunch (%) | 62.46 | 52.84 | 9.63 | 62.43 | 57.40 | 5.03 | 58.56 | 58.17 | 0.40 |
| Free or reduced lunch undefined (%) | 4.81 | 4.13 | 0.68 | 6.91 | 7.67 | –0.76 | 9.05 | 8.63 | 0.42 |

**Table 7.3—Continued**

| Student Characteristics | Year 0 | | | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference | Treatment | Control | Difference |
| ELL (%) | 12.48 | 18.02 | –5.54 | 14.14 | 21.13 | –7.00 | 15.26 | 20.87 | –5.60 |
| ELL undefined (%) | 4.81 | 4.13 | 0.68 | 6.91 | 7.67 | –0.76 | 20.04 | 18.99 | 1.06 |
| Special Ed status (%) | 10.66 | 10.59 | 0.07 | 11.37 | 11.32 | 0.05 | 6.16 | 6.38 | –0.21 |
| Special Ed status undefined (%) | 4.81 | 4.13 | 0.68 | 6.91 | 7.67 | –0.76 | 9.05 | 8.63 | 0.42 |
| Total students (no.) | 22,553 | 19,708 | | 20,683 | 19,589 | | 18,559 | 17,986 | |
| Total schools (no.) | 50 | 37 | 13 | 48 | 34 | 14 | 47 | 33 | 14 |

NOTES: Five treatment schools and five controls schools were missing in year 0. Seven treatment schools and eight control schools were missing in year 1. Eight treatment schools and nine control schools were missing in year 2.

levels when tested with a linear model with a study group indicator and cluster-adjusted standard errors for students clustered in schools.

Students from the two groups of high schools had similar performance levels but some modest differences on select Regents Exams. As Table 7.4 shows, the mean test scores in year 0 (2006–2007 school year) on the fall and spring administrations of Comprehensive English, the fall Mathematics A, Chemistry, and the spring Science–Living Environment exams were relatively similar among students enrolled in SPBP high schools and those not assigned to the program. Mean differences in test scores among SPBP and non-SPBP schools can be found on the fall administration of Mathematics B, Science–Living Environment, and Earth Science, along with the spring administration of Mathematics A, Mathematics B, and Chemistry. However, as with the elementary, middle, and K–8 statistics reported in Tables 7.1 and 7.2, these differences are not statistically significant when subjected to standard t-tests with the school-level data.

Table 7.5 shows select school-level summary statistics for high schools by treatment status and school year. It reveals very few differences in baseline student characteristics between control and treatment schools. The mean percentage of students enrolled in SPBP schools are more likely to be black than students in schools not selected to participate (43 percent to 37 percent). The opposite is true for the percentage of Hispanic students (52 percent to 57 percent). In addition, control schools have an average of about 100 more students than treatment schools. Nevertheless, these modest differences in student composition are not statistically different.

## School-Level Analysis: SPBP Effects on Progress Report Components

The SPBP bonuses were determined on the basis of the NYCDOE School Progress Reports. This section describes what they measure and the performance of the pilot study schools on the Progress Report Overall Score, the Environment Category Score, the Performance Category Score, the Progress Category Score, and Additional Credit. We

**Table 7.4**
**Select Student-Level Regents Examination Score Summary Statistics for High Schools, by Treatment Status in Baseline Year**

| | Year 0 | | |
|---|---|---|---|
| **Course and Term** | **Treatment** | **Control** | **Difference** |
| ELA | | | |
| Fall | 54.59 | 55.23 | –0.64 |
| Spring | 53.27 | 53.61 | –0.35 |
| Mathematics A | | | |
| Fall | 59.56 | 58.94 | 0.62 |
| Spring | 56.86 | 59.11 | –2.25 |
| Mathematics B | | | |
| Fall | 47.20 | 51.29 | –4.09 |
| Spring | 45.65 | 55.57 | –9.92 |
| Science–living environment | | | |
| Fall | 54.97 | 52.91 | 2.06 |
| Spring | 57.36 | 57.90 | –0.54 |
| Earth science | | | |
| Fall | 50.26 | 54.12 | –3.86 |
| Spring | 54.19 | 55.43 | –1.24 |
| Chemistry | | | |
| Fall | 49.75 | 50.14 | –0.39 |
| Spring | 53.42 | 51.71 | 1.71 |
| Number of schools | | | |
| Fall | 51 | 37 | 14 |
| Spring | 53 | 37 | 16 |
| Number of students | | | |
| Fall | 22,553 | 19,708 | 2,845 |
| Spring | 33,038 | 28,044 | 4,994 |

NOTES: Five treatment and five control schools were missing in year 0.

**Table 7.5**
**Select School-Level Summary Statistics for High Schools by Treatment Status and Year**

| | Year 0 | | | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Treat. | Cont. | Diff. | Treat. | Cont. | Diff. | Treat. | Cont. | Diff. |
| Student Characteristics (%) | | | | | | | | | |
| Female | 49.52 | 48.79 | 0.01 | 48.79 | 49.12 | –0.32 | 48.81 | 48.88 | –0.07 |
| Male | 50.48 | 51.21 | –0.73 | 51.21 | 50.88 | 0.32 | 51.19 | 51.12 | 0.07 |
| Black | 42.72 | 37.18 | 5.54 | 42.70 | 36.83 | 5.87 | 41.47 | 36.61 | 4.86 |
| Hispanic | 52.36 | 56.67 | –4.31 | 52.61 | 57.20 | –4.60 | 53.61 | 57.31 | –3.70 |
| Native Am./Am. Indian | 0.38 | 0.48 | –0.10 | 0.43 | 0.43 | 0.00 | 0.46 | 0.43 | 0.03 |
| White | 1.95 | 2.88 | –0.93 | 1.78 | 2.74 | –0.96 | 1.64 | 2.58 | –0.94 |
| Other | 2.58 | 2.79 | –0.20 | 2.49 | 2.80 | –0.31 | 2.82 | 3.06 | –0.25 |
| Free and reduced-price lunch | 70.31 | 72.22 | –1.91 | 72.87 | 72.11 | 0.77 | 71.86 | 70.71 | 1.15 |
| Average enrollment | 846 | 949 | –103 | 813 | 911 | –98 | 730 | 820 | –90 |
| Total number of schools | 50 | 37 | 13 | 48 | 34 | 14 | 47 | 33 | 14 |

NOTES: Five treatment schools and five controls schools were missing in year 0. Seven treatment schools and eight controls schools were missing in year 1. Eight treatment schools and nine controls schools were missing in year 2.

describe the performance of schools participating in SPBP each year and compare their performance with other schools. We also tested for the effects of the program on the schools for the three program years (2007–2008 to 2009–2010) and the 2006–2007 school year for baseline comparison using the original treatment and control assignments of 167 control and 232 treatment schools (i.e., using the ITT analysis).

Across all years and all the component scores, we found that SPBP had no statistically significant effect on the Progress Reports, even though these were the performance measures used to determine bonuses. This lack of effect held for elementary, middle, and high schools. We also found no differences between schools that participated in SPBP each year (regardless of random assignment) and other pilot schools.

The next section first provides a brief description of the Progress Report scores. It then summarizes the performance of the schools participating in SPBP each year and concludes with our tests of how SPBP affected Progress Report scores.

### Description of Measures

As described in Chapter Two, the Progress Report provides an overall score, with additional scores for subcomponents. Table 7.6 summarizes these various measures.[5]

Table 7.7 presents the range of overall Progress Report and component scores for schools that actually participated in SPBP and the eligible pool of schools that did not (i.e., as treated). Each year, the performance of the SPBP participants varied considerably, with some schools earning the maximum scores on each component, and others receiving very few of the total available points. Schools received particularly high scores in the 2008–2009 school year and lower scores in the 2009–2010 school year. This is also reflected in the average scores given in Table 7.8. For example, in 2008, the SPBP participant schools earned just 56.76 of the 100 possible points on average; in 2010, they earned

[5] For further descriptions, see "Criteria for Earning a Bonus: Performance Measures" in Chapter Two. See NYCDOE, 2011b, for links to further descriptions of the measures and methodology.

**Table 7.6**
**Definition of Progress Report Component Scores**

| Progress Report Component Score | Description |
|---|---|
| Overall score | Based on sum of the components below. Total of 100 possible points (plus additional credit). |
| Environment category score | Measures pre–conditions for learning, including student attendance and measures of academic expectations, communication, engagement, safety, and respect from teacher, parent, and student (middle/high school) surveys administered by NYCDOE. Total of 15 possible points. |
| Performance category score | Measures the number of students who have reached proficiency in ELA and mathematics and median proficiency levels (elementary, middle, and K–8 schools). Measures 4– and 6–Year graduation rates and weighted diploma rates (high schools). Total of 25 possible points. |
| Progress category score | Measures changes in student proficiency over the past year—average change in proficiency and percentage making progress (elementary, middle, and K–8 schools). Measures credit accumulation and Regents Exam completion and pass rates (high schools). Total of 60 possible points. |
| Additional credit | Earned when high-needs students demonstrate exemplary progress on state test scores (elementary, middle, and K–8 schools) or Regents Exam scores, credit accumulation, or graduation with Regents Diploma (high schools). Schools can earn up to 15 additional points. |

just 50 of the total points on average. In 2009, SPBP schools earned 79.03 of the possible 100 points on average. This pattern also held for the other eligible schools in the study. In fact, the score ranges and means for the SPBP participants and other schools were very similar for all years—the means for the two groups were always within about a point or two on every measure and all years. The means for the SPBP schools did tend to be slightly higher than those of the nonparticipants, but the differences were not statistically significant.[6] Moreover, these differences might not reflect the effects of the program because schools that actually participated in the program included schools that were not in the lottery and excluded schools that were assigned to the program via the lottery but declined to participate. These two groups

[6] We used a two-sample t-test to test of the differences between the two groups of schools.

**Table 7.7**
**Ranges for the Overall and Component Progress Report Scores of SPBP Participants and Other Schools, by School Year**

| | Other Schools | | SPBP Participants | |
|---|---|---|---|---|
| | Minimum | Maximum | Minimum | Maximum |
| 2007–2008 | | | | |
| Overall score | 22.9 | 100.4 | 9.6 | 101.2 |
| Category | | | | |
| Environment | 1.5 | 15.0 | 2.5 | 15.0 |
| Performance | 3.2 | 25.0 | 3.9 | 25.0 |
| Progress | 5.0 | 60.0 | 0.3 | 60.0 |
| Additional credit | 0.0 | 13.0 | 0.0 | 12.0 |
| 2008–2009 | | | | |
| Overall score | 36.0 | 108.5 | 33.5 | 106.6 |
| Category | | | | |
| Environment | 2.8 | 15 | 3.6 | 15.0 |
| Performance | 6.0 | 25 | 7.8 | 25.0 |
| Progress | 18.9 | 60 | 17.4 | 60.0 |
| Additional credit | 0.0 | 12 | 0.0 | 12.0 |
| 2009–2010 | | | | |
| Overall score | 5.1 | 93.9 | 9.3 | 105.1 |
| Category | | | | |
| Environment | 0.9 | 14.6 | 0.9 | 15.0 |
| Performance | 0.0 | 25.0 | 0.0 | 25.0 |
| Progress | 0.0 | 60.0 | 0.0 | 59.4 |
| Additional credit | 0.0 | 10.3 | 0.0 | 15.0 |

of schools could differ from other schools in ways that contributed to differences between the groups but that were not caused by SPBP participation. Schools that voted to participate might score differently than other schools on the Progress Report even in the absence of the program.

**Table 7.8**
**Means and Standard Deviations for Overall and Component Progress Report Scores of All SPBP Participants and Other Pilot Schools, by School Year**

| Progress Report Component Scores | 2007–2008[a] | | 2008–2009[b] | | 2009–2010[c] | |
|---|---|---|---|---|---|---|
| | Other | SPBP | Other | SPBP | Other | SPBP |
| Overall | 56.00 | 56.76 | 77.27 | 79.03 | 49.07 | 50.38 |
| (std. dev.) | (16.61) | (15.86) | (16.10) | (15.91) | (17.20) | (19.07) |
| Environment category | 8.20 | 8.30 | 9.60 | 9.85 | 7.81 | 7.92 |
| (std. dev.) | (2.72) | (2.63) | (2.58) | (2.52) | (2.86) | (2.76) |
| Performance category | 14.24 | 14.50 | 18.30 | 18.52 | 7.88 | 8.25 |
| (std. dev) | (4.41) | (4.05) | (4.67) | (4.46) | (4.82) | (5.42) |
| Progress category | 30.54 | 30.80 | 43.85 | 44.76 | 30.33 | 31.22 |
| (std. dev.) | (10.03) | (9.61) | (10.16) | (9.83) | (11.12) | (12.05) |
| Additional credit | 3.01 | 3.15 | 5.52 | 5.90 | 3.05 | 3.00 |
| (std. dev.) | (2.94) | (2.69) | (3.31) | (3.16) | (2.67) | (2.83) |

NOTE: Data include 408 schools that participated in the lottery or entered SPBP outside the lottery and voted to participate.

[a] 15 SPBP and 10 schools not participating in SPBP during the 2007–2008 school year are missing Progress Report Data.

[b] 17 SPBP and 18 schools not participating in SPBP during the 2008–2009 school year are missing Progress Report Data

[c] 16 SPBP and 19 schools not participating in SPBP during the 2009–2010 school year are missing Progress Report Data.

### Intent-to-Treat Analysis Results: SPBP Did Not Affect Performance on Progress Report Measures

As discussed previously, we used an ITT analysis to estimate SPBP's effects on school outcomes. In this analysis, we compared the Progress Report scores for SPBP or treatment schools (those randomly selected to participate in the program, regardless of whether school personnel actually voted in favor of participating) with control schools (those entered into the SPBP lottery but not randomly selected and invited to participate). The ITT effects were tested with standard two-sample t-test for differences in the means (Snedecor and Cochran, 1989). Tables 7.9 to 7.11 provide the effect estimates and standard errors. We provide standard errors rather than p-values because the effects are not significant, making standard errors more informative.

Tables 7.9 presents the results for all schools. Tables 7.10 and 7.11 present the results for elementary, middle, and K–8–high schools respectively. Schools were classified according to the school classification presented in Chapter Two, with middle–high schools in both the elementary, middle, K–8 school-level and the high school-level samples. There were 324 elementary, middle schools, and K–8 schools in the data and 97 high schools.

We found no statistically significant SPBP effects in any year. The group means were nearly identical for many component scores in most years. This lack of effects held for all levels of schooling. We also estimated effects for 2006–2007 as a test for possible baseline differences due to an imbalance in the randomization. Again there were no significant differences, except for positive a difference between SPBP and other high school performance scores, and the estimated differences were generally small and qualitatively very similar to the estimated effects for the intervention years.

## Student-Level Analysis: SPBP's Effect on Student Test Scores

The purpose of this analysis was to estimate SPBP's effect on student test scores using assessment results for students in grades 3 to 12 during

**Table 7.9**
**Estimated SPBP Effects on Overall and Component Progress Report Scores, by School Year**

| Progress Report Component Score | School Year | | |
|---|---|---|---|
| | 2007–2008[a] | 2008–2009[b] | 2009–2010[b] |
| Overall | −0.39 | 0.11 | −1.32 |
| (std. err.) | (1.65) | (1.63) | (1.86) |
| Categories | | | |
| Environment | −0.06 | 0.14 | −0.12 |
| (std. err.) | (0.27) | (0.26) | (0.29) |
| Performance | 0.15 | 0.00 | −0.31 |
| (std. err.) | (0.42) | (0.47) | (0.52) |
| Progress | −0.28 | 0.01 | −0.60 |
| (std. err.) | (1.00) | (1.02) | (1.19) |
| Additional credit | −0.20 | −0.03 | −0.29 |
| (std. err.) | (0.29) | (0.33) | (0.28) |

[a] High school data for the 2007–2008 school year include only 87 schools with Progress Reports—six control and four SPBP schools were missing data.

[b] Elementary and middle school data for the 2008–2009 and 2009–2010 school years include only 306 schools with Progress Reports—nine SPBP and five control schools were missing data, and four SPBP schools had closed; high school data for the 2008–2009 and 2009–2010 school years include only 87 schools with Progress Reports—four control and six SPBP schools were missing data.

the four school years of interest. Across all three program years for elementary, middle, and K–8 levels, we found that SPBP did not improve average student test scores in mathematics and ELA. Average scores for student enrolled in elementary, middle, and K–8 schools invited to participate in SPBP were lower than those for students enrolled in control schools for both subjects and all years, but the differences were small and were statistically significant only for mathematics in the last year. Similarly, across two years for high schools, we found no effect on student test scores in any of the Regents subjects: Comprehensive English, Mathematics A, Mathematics B, Science–Living Environment, Earth

**Table 7.10**
**Estimated SPBP Effects on Overall and Component Progress Report Scores for Elementary and Middle Schools, by School Year**

| Progress Report Component Score | School Year | | |
|---|---|---|---|
| | 2007–2008 | 2008–2009[a] | 2009–2010[a] |
| Overall Score | 0.21 | –1.14 | –2.62 |
| (std. err.) | (1.82) | (1.42) | (2.02) |
| Categories | | | |
| Environment | 0.00 | 0.11 | –0.17 |
| (std. err.) | (0.29) | (0.28) | (0.32) |
| Performance | 0.33 | –0.04 | –0.35 |
| (std. err.) | (0.46) | (0.44) | (0.44) |
| Progress | 0.10 | –0.95 | –1.64 |
| (std. err.) | (1.14) | (0.97) | (1.37) |
| Additional credit | –0.22 | –0.25 | –0.46 |
| (std. err.) | (0.32) | (0.35) | (0.32) |

[a] Data for the 2008–2009 and 2009–2010 school years include only 306 schools with Progress Reports—nine SPBP and five control school were missing data, and four SPBP schools had closed.

Science, and Chemistry. The next section begins with a brief description of our analytic approach, followed by a summary of findings.

### Analytic Approach

As discussed previously, we used an ITT analysis to estimate how SPBP might affect student achievement test scores. We produced estimates for high schools separately from those for elementary, middle, and K–8 schools because Regents Exams differ from the exams administered to students in grades 3 through 8. The content assessed on Regents Exams is more narrowly focused than that for the lower grades. Students can also select when to take Regents Exams and can sit for an exam multiple times. Although we restricted our analysis sample to scores from

**Table 7.11**
**Estimated SPBP Effects on Overall and Component Progress Report Scores for High Schools, by School Year**

| Progress Report Component Score | School Year | | |
|---|---|---|---|
| | 2007–2008[a] | 2008–2009[b] | 2009–2010[b] |
| Overall Score | –2.68 | 2.64 | 4.63 |
| (std. err.) | (3.85) | (3.69) | (3.44) |
| Categories | | | |
| Environment | –0.29 | 0.12 | 0.22 |
| (std. err.) | (0.66) | (0.58) | (0.54) |
| Performance | –0.56 | –0.35 | 0.51 |
| (std. err.) | (1.01) | (0.95) | (1.09) |
| Progress | –1.71 | 2.34 | 3.65 |
| (std. err.) | (2.04) | (2.16) | (1.99) |
| Additional Credit | –0.12 | 0.53 | 0.25 |
| (std. err.) | (0.67) | (0.71) | (0.59) |

[a] Data for the 2007–2008 school year include only 87 schools with Progress Reports—six control and four SPBP schools were missing data.

[b] Data for the 2008–2009 and 2009–2010 school years include only 87 schools with Progress Reports—four control and six SPBP schools were missing data.

the first time a student took a specific examination,[7] the lack of uniformity on assessments and the assessment process across the different grade levels could have produced systematic differences in the SPBP

[7] The common interpretation of the SPBP effect on achievement is the average of the difference between a student's achievement had the student been enrolled in a school participating in SPBP and the achievement had the student been enrolled in a school not participating in SPBP. Because high school students have some flexibility in taking the Regents Exams, SPBP could affect which students take an exam and when; for students who take the exam, it could also affect performance. Our estimate of the SPBP effect combines the effects on who takes each exam and how they perform. It describes how much, on average, we might expect student achievement on the Regents Exams to change because of a school's participation in SPBP. Our high school analyses did not estimate how much an individual student's performance would change as a result of the school's participation in the program. For elementary and middle school grades, the tests are mandatory, and we interpreted the SPBP effect for schools with those grades as the average effect on the achievement of an individual student.

effects on student test scores for reasons unrelated to the incentive pay program itself. Also, high school operations differ from other grade levels, making it valuable to consider the program effects separately. Therefore, schools that enroll students in any of grades 3 to 8 and any of grades 9 to 12 are included in separate analyses with the appropriate subsets of their students used in each analysis.

Both the elementary, middle, and K–8 school and the high school analyses implemented the ITT analysis using ordinary least squares methods with cluster adjusted standard errors at the school level to correct for possible similarities among test scores from students who attended the same schools, which could diminish the effective sample size of the observations and, consequently, the precision of the estimates.

We present estimates from the simple ITT approach in this chapter and the adjusted ITT estimates with and without student- and school-level controls in Appendix D. Select model specifications included student and school-level covariates from the school year just prior to year 1 of SPBP. Including student- and school-level controls can help reduce statistical errors in the group average. These controls provide the analyses with more precise estimates and, consequently, greater ability to determine whether SPBP had an effect on student test scores. More specifically, select model specifications include the following:

- student-level controls
  - binary indicators for free- and reduced-price lunch status
  - ELL status
  - special education status
  - race and/or ethnicity (American Indian or Alaskan Native, Asian or Pacific Islander, black, Hispanic or Latino, other, and white
  - gender
  - whether a student was retained (i.e., enrolled in same grade in both year t and year t-1).
- school-level controls
  - the number of students enrolled in the school

- the percentage of students receiving free and reduced-price lunch
- the percentage of students classified as ELL
- the percentage of students receiving special education services
- the percentage of students by race and/or ethnicity
- the percentage of students who are male
- binary indicators for school type (i.e., elementary, middle, or K–8).

For elementary, middle, and K–8 school students, we also estimated a series of models that contained all these student- and school-level variables and a student's prior ELA test scores for models using ELA scores as the dependent variable and a student's prior mathematics test scores for models using mathematics scores as the outcome of interest. A student's prior test score was measured in the most recent year outside the time he or she attended a school during the evaluation. For example, a student who enrolled in a study school all three years of the program (i.e., the 2007–2008, 2008–2009, and 2009–2010 school years), the prior achievement would be from the 2006–2007 school year; for a student who transferred to an SPBP school at the start of year 3 (the 2009–2010), prior achievement would be measured in the 2008–2009 school year. Even though the current assessment was first used in the 2005–2006 school year, prior test scores were restricted to scores from the 2006–2007 school year.

### Sensitivity Analyses

We conducted various robustness checks to explore whether estimates of SPBP effect were sensitive to the specific scales we report here or to the model specification. We considered an alternative test score transformation for every grade level, subject, and year. To calculate this alternative transformation, we first partitioned the student scores into 20 bins based on equal length intervals of the students' prior year assessment scores. Within each bin, we created alternative z-score transformations using the mean ($m_b$) and standard deviation ($s_b$) for students in the bin, $b$=1, . . . , 20:

$$Z_{alt} = \frac{y - m_b}{s_b} \text{ for a student in bin } b.$$

We then reestimated our models using these alternatively transformed scores. We also considered models with alternative sets of covariates. We fit models with and without prior test scores because many students do not have any prior scores (e.g., they were enrolled in grades that were not tested in years before enrolling in a school in the study or they transferred into the New York City school system). We also ran models that contain prior test scores expressed as the quadratic or cubic of a student's test score in the most recent year prior to a school participating in SPBP. The results proved to be very similar to those reported below.

We also tested for interactions between school size and SPBP. As noted in Goodman and Turner (2011), larger schools might be more susceptible to "free riders" or staff feeling personally responsible for achieving the bonus. Moreover, Goodman and Turner report such interactions using school-level analyses.

#### Relationship Between Distribution Plans and Student Outcomes

We also examined whether aspects of the bonus distribution plans that SPBP elementary, middle, and K–8 schools adopted were associated with better outcomes in the 2009–2010 school year. To operationalize different bonus distribution plans between schools, we calculated the Gini coefficient using award distribution information for all school employees in the school, for teachers only, and for teachers and coaches only. We also calculated the proportion of bonus awards at the school that did not equal the modal award for the school. Finally, we created an indicator denoting whether a school differentiated bonus awards on the basis of individual performance. Estimates from these analyses are reported in Appendix D.

## Student Test Score Findings

This section presents our results on how SPBP affected student test scores during its three years of implementation. The first set of results focuses on the elementary, middle, and K–8 school sample for years 1 to 3. The second reports findings for the high school sample in years 1 and 2.

### SPBP Did not Improve Average Mathematics or ELA Test Scores for Elementary, Middle, and K–8 Students

Table 7.12 shows our results on the effects of SPBP on student test scores in mathematics and ELA for students enrolled in elementary, middle, and K–8 level schools. The top half of the table contains estimates using the original, unstandardized scale scores as the student performance measure. The bottom half reports estimates using the standardized scale scores as the outcome variable. Each set of analyses contains estimates with and without controls for student- and school-level characteristics.[8]

The estimates show that the average mathematics and ELA test scores of students from schools randomly chosen for an invitation to participate in SPBP were lower than those of students from control schools during years 1, 2, and 3. The magnitudes of the estimates, however, were very small—on the order of about –0.97 to –2.13 scale score points in mathematics and –0.69 and –0.91 scale score points in ELA on a test scale that typically ranges 300 points at each grade level.[9] Further, the differences were statistically significant only for mathematics in year 3. Conducting multiple statistical tests from the same experiment, such as testing the SBPB effects for two subjects in each of

[8] The estimates reported in Tables 7.12 through 7.14 (and Appendix F) do not take into account a student's prior test performance when the controls are added for student- and school-level characteristics. Analyses controlling for a students' prior test performance yielded similar results but were not reported in these tables because of the number of students lacking prior test scores, which results in loss of student observations.

[9] The State of New York mathematics and ELA assessments have two key features in common across grades 3 through 8: a score of 650 is the minimum score for proficiency under NCLB (minimum Level 3 score), and all have a standard deviation of 40.

three years, could result in one or more spurious statistically significant effects even if the program had no true effects. When we adjusted for the multiple comparisons using the approach of Benjamini and Hochberg (1985), the year 3 mathematics effects were no longer significant, suggesting that, because of chance differences in the schools assigned to the SPBP and control groups, the effects we observed would not be uncommon. Moreover, ELA scores for SPBP schools were lower on average than the corresponding scores for control schools in year 1 even before the schools participated in the program, further suggesting that the observed differences were due to chance assignment, not the effects of the program. The results provided in the bottom panel of Table 7.12 are qualitatively similar, suggesting our results are robust to the particular specification of the model and scaling of the test scores.

### SPBP Had No Discernable Average Effect on Regents Exam Scores for High School Students in Years 1 and 2

Tables 7.13 and 7.14 show the average effects of SPBP on Regent Exam scores for ELA, Mathematics A, Mathematics B, Science–Living Environment, Earth Science, and Chemistry. Table 7.13 provides results using scales scores as the student performance measure and then Table 7.14 offers results using the standardized scores as the dependent variable. The tables report effects for both the fall and spring administrations. For each test and test administration, the tables also report the number of schools included in the analysis. These numbers vary because not all schools administered a specific Regents Exam or not all students were participating in Regents testing for the first time and/or completed the exam during each administration.

As shown in Table 7.13 and Table 7.14, the results of the ITT analysis indicate that SPBP had no discernable effect on Regents Exam scores. Across all tests and test administrations, the estimated effects are very small, many with effect sizes less than 0.05 in absolute value. None of the effects are significant.

**Table 7.12**
**SPBP Effects on Test Scores for Elementary, Middle, or K–8 Schools, by Subject and School Year (ITT Estimates)**

| | Year 1 | | Year 2 | | Year 3 | | Years 1, 2, and 3 | |
|---|---|---|---|---|---|---|---|---|
| | No Controls | Controls | No Controls | Controls | No Controls | Controls | No Controls | Controls |
| Scale Scores | | | | | | | | |
| Mathematics | –1.20 | –0.97 | –1.63 | –1.59 | –2.13* | –1.93* | –1.65 | –1.57 |
| (std. err.) | (1.08) | (0.89) | (1.05) | (0.90) | (1.07) | (0.96) | (1.01) | (0.86) |
| Observations | 111,014 | 111,014 | 108,093 | 107,953 | 108,555 | 107,976 | 327,662 | 326,943 |
| $R^2$ | 0.10 | 0.23 | 0.09 | 0.20 | 0.09 | 0.22 | 0.10 | 0.22 |
| ELA | — | — | –0.69 | –0.80 | –0.90 | –0.91 | — | — |
| (std. err.) | — | — | (0.78) | (0.65) | (0.66) | (0.58) | — | — |
| Observations | — | — | 105,483 | 105,330 | 105,303 | 105,269 | — | — |
| $R^2$ | — | — | 0.04 | 0.18 | 0.08 | 0.24 | — | — |
| Student and school controls | | √ | | √ | | √ | | √ |
| Standardized Scores | | | | | | | | |
| Mathematics | –0.03 | –0.02 | –0.05 | –0.04 | –0.06* | –0.06* | –0.05 | –0.04 |
| (std. err.) | (0.03) | (0.02) | (0.03) | (0.02) | (0.03) | (0.03) | (0.03) | (0.02) |
| Observations | 111,014 | 111,014 | 108,093 | 107,953 | 108,555 | 107,976 | 327,662 | 326,943 |
| $R^2$ | 0.00 | 0.15 | 0.01 | 0.13 | 0.01 | 0.15 | 0.01 | 0.14 |

**Table 7.12—Continued**

| | Year 1 | | Year 2 | | Year 3 | | Years 1, 2, and 3 | |
|---|---|---|---|---|---|---|---|---|
| | No Controls | Controls | No Controls | Controls | No Controls | Controls | No Controls | Controls |
| ELA | — | — | –0.02 | –0.03 | –0.03 | –0.03 | — | — |
| (std. err.) | — | — | (0.02) | (0.02) | (0.02) | (0.02) | — | — |
| Observations | — | — | 105,483 | 105,330 | 105,303 | 105,269 | — | — |
| $R^2$ | — | — | 0.00 | 0.15 | 0.01 | 0.18 | — | — |
| Student and school controls | | √ | | √ | | √ | | √ |
| Number of schools | | | | | | | | |
| Treatment | 191 | | 187 | | 187 | | 191 | |
| Control | 133 | | 133 | | 132 | | 133 | |

NOTES: *, **, *** Estimates are statistically significant from zero at the 5%, 1%, and 0.1% levels, respectively. Student-level controls include gender, race and/or ethnicity, ELL status, special education status, free or reduced-price lunch status, retention flag, and grade level. School-level controls include male students (%), students by race and/or ethnicity (%), free and reduced-price lunch students (%), school size, and school level. The referant categories are grade 8, white, female, white (%), female (%), K–8 schools, and year 3. Four of the 324 in the elementary, middle, and K–8 sample were excluded from the analysis because they did not enroll students in at least one tested grade (grades 3–8). Standard errors appear in parentheses. Year 1 and three-year combined results for ELA are excluded because testing occurred early in the school year, before effects could occur.

### No Statistically Significant Relationship Was Found Between Student Achievement and Bonus Award Plans

For elementary, middle, and K–8 school students, we estimated the relationship between achievement levels and the school's implementation of the bonus program, as measured by the bonus differentiation in its 2010 distribution plan. We used the Gini coefficient for all staff, teachers, teachers, and coaches; the proportion of staff not receiving the modal award for the school; and whether or not the school reported using individual performance as part of the basis for determining awards. For each of these measures of award differentiation, we estimated the change in student mathematics scores and ELA scores associated with differences in the measures. The results presented in Table 5 in Appendix D show very small effects for the relationship between outcomes and the Gini coefficient or classification of schools on the basis of the survey measures. The largest relationship is between the proportion of staff slated not to receive the school's modal award and student achievement, but none of the effects are statistically significant. The small number of schools with significant differentiation limited our ability to model these relationships.

## Summary and Discussion

Overall, we found that SPBP had no statistically significant effects on school Progress Report scores. Across all years and all component scores for the Progress Reports (environment, performance, progress, and additional credit), we found no statistically significant differences between scores of SPBP treatment and control schools or between schools that participated in SPBP each year (regardless of random assignment) and other eligible schools. This lack of effects held true for elementary, middle, and high schools.

In addition, we found that SPBP did not improve student test scores in mathematics and ELA during the first three years of the program at the elementary, middle, and K–8 levels and the first two years of the program at the high school level. We also found that SPBP did not improve student test scores after adjusting estimates for schools

**Table 7.13**
**SSPBP Effects on Regents Exams Scores for High Schools, by Subject and School Year: Scale Scores (ITT Estimates)**

| | Fall Test Administration | | Spring Test Administration | | | |
|---|---|---|---|---|---|---|
| | Fall, Year 2 | | Spring, Year 1 | | Spring, Year 2 | |
| | No Controls | Controls | No Controls | Controls | No Controls | Controls |
| ELA | 0.62 | 0.72 | 0.39 | –0.56 | 2.23 | 1.72 |
| (std. err.) | (2.01) | (2.14) | (1.79) | (1.63) | (1.52) | (1.37) |
| Student Observations | 8,234 | 8,234 | 10,276 | 10,276 | 9,345 | 9,345 |
| School Observations | 82 | 82 | 84 | 84 | 82 | 82 |
| $R^2$ | 0.00 | 0.17 | 0.00 | 0.15 | 0.00 | 0.19 |
| Mathematics A | 0.26 | 0.48 | –0.16 | 0.54 | — | — |
| (std. err.) | (0.81) | (0.66) | (1.07) | (0.97) | — | — |
| Student Observations | 8,469 | 8,469 | 11,907 | 11,907 | — | — |
| School Observations | 82 | 82 | 82 | 82 | — | — |
| $R^2$ | 0.00 | 0.09 | 0.00 | 0.10 | — | — |
| Mathematics B | –0.66 | –0.88 | –2.32 | –0.19 | –2.72 | 0.97 |
| (std. err.) | (3.13) | (2.38) | (2.66) | (2.20) | (2.72) | (2.28) |
| Student Observations | 1,299 | 1,299 | 1,914 | 1,914 | 1,813 | 1,813 |
| School Observations | 62 | 62 | 66 | 66 | 72 | 72 |
| $R^2$ | 0.00 | 0.10 | 0.00 | 0.11 | 0.01 | 0.11 |

**Table 7.13—Continued**

| | Fall Test Administration | | Spring Test Administration | | | |
|---|---|---|---|---|---|---|
| | Fall, Year 2 | | Spring, Year 1 | | Spring, Year 2 | |
| | No Controls | Controls | No Controls | Controls | No Controls | Controls |
| Integrated Algebra | 0.65 | 1.36 | 1.14 | 1.58 | –1.14 | –0.28 |
| (std. err.) | (1.09) | (0.78) | (1.06) | (1.11) | (1.04) | (1.03) |
| Student Observations | 7,170 | 7,170 | 10,447 | 10,447 | 16,062 | 16,062 |
| School Observations | 81 | 81 | 80 | 80 | 82 | 82 |
| $R^2$ | 0.00 | 0.08 | 0.00 | 0.09 | 0.00 | 0.09 |
| Science–Living Environment | 0.08 | –0.42 | 0.15 | 0.27 | –0.36 | 0.30 |
| (std. err.) | (1.21) | (1.20) | (1.07) | (1.08) | (1.12) | (1.17) |
| Student Observations | 7,903 | 7,903 | 11,953 | 11,953 | 12,540 | 12,540 |
| School Observations | 82 | 82 | 81 | 81 | 81 | 81 |
| $R^2$ | 0.00 | 0.07 | 0.00 | 0.09 | 0.00 | 0.13 |
| Earth Science | –2.38 | –2.03 | –2.70 | –2.84 | –1.36 | –1.89 |
| (std. err.) | (2.82) | (2.22) | (1.85) | (1.87) | (2.12) | (2.11) |
| Student Observations | 1,490 | 1,490 | 5,856 | 5,856 | 6,197 | 6,197 |
| School Observations | 50 | 50 | 68 | 68 | 68 | 68 |
| $R^2$ | 0.01 | 0.10 | 0.01 | 0.11 | 0.00 | 0.10 |

**Table 7.13—Continued**

| | Fall Test Administration | | Spring Test Administration | | | |
|---|---|---|---|---|---|---|
| | Fall, Year 2 | | Spring, Year 1 | | Spring, Year 2 | |
| | No Controls | Controls | No Controls | Controls | No Controls | Controls |
| Chemistry | –3.76 | –3.56 | 0.01 | 0.13 | 1.62 | 3.40 |
| (std. err.) | (5.74) | (3.10) | (2.02) | (1.88) | (2.38) | (2.14) |
| Student Observations | 316 | 316 | 2,869 | 2,869 | 2,705 | 2,705 |
| School Observations | 31 | 31 | 62 | 62 | 57 | 57 |
| $R^2$ | 0.02 | 0.41 | 0.00 | 0.08 | 0.00 | 0.12 |
| Student and school controls | | √ | | √ | | √ |

NOTES: *, **, *** Estimates statistically significant from zero at the 5%, 1%, and 0.1% levels, respectively. Student-level controls include gender, race and/or ethnicity, free/reduced lunch status, and indicators if gender or race and/or ethnicity was undefined. School-level controls include the percentage of male students, percentage of students by race and/or ethnicity, percentage of free/reduced price lunch students, school level and school size. The number of schools containing at least one student with valid test score information ranges between 31 (Chemistry, fall 2008) and 84 (ELA, spring 2008). Table does not contain estimates for Mathematics–A in spring 2009 because of insufficient number of valid test scores. Standard errors appear in parentheses.

**Table 7.14**
**SPBP Effects on Regents Exams Scores for High Schools, by Subject and School Year: Standardized Scores (ITT Estimates)**

| | Fall Test Administration | | Spring Test Administration | | | |
|---|---|---|---|---|---|---|
| | Fall, Year 2 | | Spring, Year 1 | | Spring, Year 2 | |
| | No Controls | Controls | No Controls | Controls | No Controls | Controls |
| ELA | 0.04 | 0.04 | 0.02 | -0.03 | 0.13 | 0.09 |
| (std. err.) | (0.11) | (0.11) | (0.09) | (0.08) | (0.07) | (0.07) |
| Student Observations | 8,234 | 8,234 | 10,276 | 10,276 | 9,345 | 9,345 |
| School Observations | 82 | 82 | 84 | 84 | 82 | 82 |
| $R^2$ | 0.00 | 0.17 | 0.00 | 0.14 | 0.00 | 0.19 |
| Mathematics A | 0.02 | 0.04 | -0.01 | 0.04 | … | … |
| (std. err.) | (0.07) | (0.05) | (0.08) | (0.07) | … | … |
| Student Observations | 8,469 | 8,469 | 11,907 | 11,907 | … | … |
| School Observations | 82 | 82 | 83 | 83 | … | … |
| $R^2$ | 0.00 | 0.08 | 0.00 | 0.09 | … | … |
| Mathematics B | -0.04 | -0.05 | -0.13 | 0.01 | -0.15 | 0.05 |
| (std. err.) | (0.19) | (0.14) | (0.15) | (0.13) | (0.15) | (0.13) |
| Student Observations | 1,299 | 1,299 | 1,914 | 1,914 | 1,813 | 1,813 |
| School Observations | 62 | 62 | 66 | 66 | 72 | 72 |
| $R^2$ | 0.00 | 0.08 | 0.00 | 0.07 | 0.01 | 0.08 |

**Table 7.14—Continued**

| | Fall Test Administration | | Spring Test Administration | | | |
|---|---|---|---|---|---|---|
| | Fall, Year 2 | | Spring, Year 1 | | Spring, Year 2 | |
| | No Controls | Controls | No Controls | Controls | No Controls | Controls |
| Integrated Algebra | 0.05 | 0.11 | 0.08 | 0.12 | -0.08 | 0.22 |
| (std. err.) | (0.09) | (0.06) | (0.08) | (0.08) | (0.08) | (0.07) |
| Student Observations | 7,170 | 7,170 | 10,447 | 10,447 | 13,855 | 13,855 |
| School Observations | 81 | 81 | 80 | 80 | 82 | 82 |
| $R^2$ | 0.00 | 0.08 | 0.00 | 0.09 | 0.00 | 0.09 |
| Science–Living Environment | 0.01 | -0.03 | 0.01 | 0.02 | -0.03 | 0.02 |
| (std. err.) | (0.09) | (0.09) | (0.08) | (0.08) | (0.08) | (0.09) |
| Student Observations | 7,903 | 7,903 | 11,953 | 11,953 | 12,540 | 12,540 |
| School Observations | 82 | 82 | 81 | 81 | 81 | 81 |
| $R^2$ | 0.00 | 0.07 | 0.00 | 0.09 | 0.00 | 0.12 |
| Earth Science | -0.15 | -0.13 | -0.16 | -0.17 | -0.08 | -0.12 |
| (std. err.) | (0.18) | (0.15) | (0.11) | (0.11) | (0.13) | (0.13) |
| Student Observations | 1,490 | 1,490 | 5,856 | 5,856 | 6,197 | 6,197 |
| School Observations | 50 | 50 | 68 | 68 | 68 | 68 |
| $R^2$ | 0.01 | 0.09 | 0.01 | 0.10 | 0.00 | 0.09 |

**Table 7.14—Continued**

| | Fall Test Administration | | Spring Test Administration | | | |
|---|---|---|---|---|---|---|
| | Fall, Year 2 | | Spring, Year 1 | | Spring, Year 2 | |
| | No Controls | Controls | No Controls | Controls | No Controls | Controls |
| Chemistry | -0.30 | -0.28 | 0.00 | 0.01 | 0.13 | 0.26 |
| (std. err.) | (0.45) | (0.24) | (0.17) | (0.16) | (0.18) | (0.17) |
| Student Observations | 316 | 316 | 2,869 | 2,869 | 2,705 | 2,705 |
| School Observations | 32 | 32 | 62 | 62 | 57 | 57 |
| $R^2$ | 0.09 | 0.39 | 0.00 | 0.05 | 0.00 | 0.11 |
| Student and school controls | | √ | | √ | | √ |

NOTES: *, **, and *** Estimates are statistically significant from zero at the 5%, 1%, and 0.1% levels, respectively. Student-level controls include gender, race and/or ethnicity, free or reduced-price lunch status; indicators of gender or race and/or ethnicity were undefined. School-level controls include male students (%), students by race and/or ethnicity (%), free and reduced-price lunch students (%), school level, and school size. The number of schools containing at least one student with valid test score information ranged between 31 (chemistry, fall 2008) and 84 (ELA, spring 2008). Table does not contain estimates for Mathematics A in spring 2009 because of insufficient valid test scores. Standard errors appear in parentheses.

assigned to the SPBP intervention that declined participation. The same held across a number of robustness checks, including different mixtures of controls for student- and the school-level covariates, whether the dependent variable was a raw or standardized score, and the methodological approach to account for the clustered structure of the data.

As noted earlier, during the SPBP program and our evaluation, New York state determined that the type of questions on its test were too easy for teachers to predict, suggesting that the resulting scores might be inaccurate measures of student knowledge and proficiency levels. This does not invalidate our estimate of the SPBP effect on the tests. Both SPBP and the control groups were measured on the same tests with the same limitations, so differences between groups could not be due to the test alone. Similarly, any assessment revisions and proficiency standard changes in year 3 (2009–2010 school year) did not invalidate our test of SPBP's effect on scores from those assessments because the changes applied to both groups. However, the problems with the assessments could have implications for the program's potential to affect outcomes. In years 1 and 2, teachers' efforts to improve learning might not have been captured by tests that teachers in both SPBP and control groups had corrupted by teaching to item types or very narrowly tested material.

More generally, the problems with the New York assessments point to the limitations of using test scores as outcome measures. They are imperfect measures of student knowledge and skills and might not capture valuable effects of the program. Concerns have been raised in both the media and academic literature about unintended consequences of high-stakes testing programs (e.g., Koretz, 2002). However, test scores are the best available uniform measure for tracking student outcomes, in spite of their limitations (McCombs, Kirby, and Mariano, 2009). Moreover, school systems are interested in improving accountability test scores, and one requirement of any effective performance-pay system would be that it lead to improvements in the performance measures beyond what would have occurred were it not in place. Consequently, even though the state assessments were limited,

our tests of the effects of the program on these measures were essential to our evaluation.

CHAPTER EIGHT

# Teacher Attitudes and Behaviors in SPBP and Control Schools

In this chapter, we assess how SPBP affected self-reported attitudes and behaviors by comparing the responses of a sample of teachers in all SPBP-assigned schools with the responses of teachers in eligible schools that were not randomly selected. Drawing on results from surveys administered in spring 2010, this analysis helps us determine whether SPBP had an effect on teachers that might have been the mediator or source of effects on student achievement.

As noted at the outset, the leaders who designed SPBP set out several hypotheses about how the program would improve student outcomes:

- The possibility of receiving a share of the financial bonus would motivate individual staff to improve their efforts and practices to maximize chances of achieving that bonus, which would improve practices and ultimately improve student achievement.
- Winning a financial bonus would enhance staff morale, which would improve practices and student achievement.
- Providing a bonus tied to schoolwide performance would incentivize all staff to work together to maximize chances of achieving that bonus, which would improve student achievement either directly or indirectly as a result of improved school and classroom practices.
- Participating in the bonus program might motivate staff to seek and sustain employment at these schools, which could in turn

provide greater continuity and support for school improvement efforts and ultimately improve student achievement.

This chapter helps test these hypotheses by investigating whether indicators of these intermediate outcomes—instructional practices, morale, collaboration, and staff retention—differ between program and control schools. This analysis also helps explain the results described in the previous chapter, on student achievement. As we describe, given the lack of differences in teacher attitudes and behaviors, it is not surprising that we found that SPBP did not affect student achievement. One would not expect to see higher student test results if nothing different was occurring in terms of instruction, teacher effort, collaboration, or retention in SPBP schools from in control schools.

The research question answered in this chapter is: How did SPBP affect teachers' self-reported attitudes and behaviors?

## Methodology

Chapter Three provides details about the survey administration. As a reminder, we chose to conduct an ITT analysis to examine how SPBP affected teacher outcomes. We compared the attitudes and behavior of teachers from schools randomly assigned to the SPBP group with those from schools assigned to control regardless of the school's actual SPBP participation. This provided valid tests of the effect of the intervention without biases from differences between schools that do or do not opt to participate in the program when given a choice.

### Sample

In total, 228 treatment schools and 162 control schools were included in the teacher survey (Table 8.1).[1] We received responses from teachers in 227 treatment schools (roughly 100 percent) and 160 control schools (99 percent). The teacher-level response rate was 59 percent

[1] As discussed in Chapter Three, four SPBP-assigned were not included in the survey sampling frame because the schools had closed, and six control-group schools were not included because of errors in survey administration.

**Table 8.1**
**School- and Teacher-Level Response Rates for SPBP Teacher Survey (2010)**

| | School | | Teacher | | | |
|---|---|---|---|---|---|---|
| | | | Treatment | | Control | |
| | Treatment | Control | Tested | Untested | Tested | Untested |
| Total fielded | 228 | 162 | 896 | 670 | 637 | 475 |
| Total responded | 227 | 160 | 544 | 379 | 351 | 258 |
| Percent | | | 61 | 57 | 55 | 54 |
| Percent responded | 100 | 99 | 59 | | 55 | |

for the treatment schools and 55 percent for the control schools. We examined potential differences between responding teachers and nonresponders based on the limited data we had on teacher characteristics, which only included whether a teacher taught a tested or nontested subject or grade (i.e., mathematics or ELA) or not. We did not find significant differences in the composition of respondents and nonrespondents based on teachers' tested or nontested status. As noted in Chapter Three, to account for possible differential response rates between SPBP and control-group teachers, we controlled for the teachers and school variables described below in the models that test for program effects on survey responses.

### Survey Measures

We created scales from the survey responses corresponding to selected attitudes, instructional practices, professional development, and school environment factors. In some cases, we relied on individual items; in others, we created scales by combining responses across multiple items. To create these composite measures, we reviewed each of the survey questions, computed descriptive statistics for all item-level responses (including examining full distributions of responses to each item), examined the internal consistency (reliability coefficient) among items designed to measure common constructs, and conducted exploratory factor analyses where appropriate.

In total, we created 14 scales (grouped topically in Tables 8.2 through 8.7). Scale scores were computed by averaging the item scores across the responded items in the scale. These tables also present the number of items in each scale and the internal consistency (coefficient Alpha) values for all scales.

**Analysis**

We used a mixed-effects two-level hierarchical model to examine whether teachers in the treatment schools differed significantly from their counterparts in the control schools on the survey scales. Additional details on this model are in Appendix E. The coefficient on the SPBP indicator equals the estimated difference in the scale scores between teachers in the treatment and control schools, after controlling for school grade levels and demographic variables, and estimates the SPBP effect on the scales or items.

For individual items, we applied an ologit (ordinal logit) regression model for items with more than two choices (e.g., from strongly disagree to strongly agree) and a logit regression model for items with two choices (e.g., yes or no), with the same set of school-level covariates and adjustment for clustered data (Long, 2006).

If we were only interested in the SPBP effect on a specific outcome measure (i.e., a scale or an item), p-values smaller than 0.05 would indicate a statistically significant treatment effect on that outcome measure. However, we were interested in testing the overall program effect across many outcomes. In total, we tested the program effect on 37 outcome measures, including 14 scales and 23 individual items. The likelihood of finding a spurious significant treatment effect when the true effects are all null increases with the number of tests conducted. Therefore, caution is warranted when interpreting the significance of treatment coefficients and making inferences about the program effect. In this chapter, we report both the traditional p-values and the significance of effects after adjusting for the increased risk of finding one or more spurious effects. Although three outcome measures that showed a significant treatment effect based on a significance level of 0.05 suggest differences in these outcomes between SPBP and control-group teachers, only one of these had a significant treatment coefficient after

**Table 8.2**
**SPBP Teacher Survey Scale Description: Instructional Practices**

| Construct | Description |
|---|---|
| **Instructional practices scale:** | Never = 1<br>Every few months = 2<br>Once or twice a month = 3<br>Once or twice a week = 4<br>Almost daily = 5 |
| Test-score-driven instruction<br>Items: 9<br>Reliability coeff.: 0.85 | I use test score data to —<br>– identify individual students who need additional assistance<br>– set learning goals for individual students<br>– tailor instruction to individual students' needs<br>– develop recommendations for tutoring or other educational support services<br>– assign or reassign students to groups within my class<br>– identify topics requiring more or less emphasis in instruction<br>– encourage parent involvement in student learning<br>– identify areas where I need to strengthen my content knowledge or skills<br>– reflect on and discuss teaching and learning with my inquiry team or other teachers, coaches, etc. |
| Student-directed instruction<br>Items: 4<br>Reliability coeff.: 0.66 | How often do you engage in the following activities as part of your instructional practice?<br>– Plan different assignments or lessons for groups of students based on their performance on assignments or assessments<br>– Re–teach topics because student performance on assignments or assessments did not meet high standards<br>– Conduct preassessment to find out what students know about a topic before teaching it<br>– For those students who are having academic problems, I try to make direct contact with their parents |
| Standards-aligned instruction<br>Items: 2<br>Reliability coeff.: 0.54 | How often do you engage in the following activities as part of your instructional practice?<br>– Analyze students' work to identify the curricular standards or content that students have or have not yet mastered<br>– Design my classroom lessons to be aligned with specific curricular standards |

**Table 8.3**
**SPBP Teacher Survey Scale Description: Professional Development**

| Construct | Description |
|---|---|
| **Professional development scale:** | None = 1<br>5 hours = 2<br>6–20 hours = 3<br>21–40 hours = 4<br>More than 40 hours = 5 |
| On instruction<br>Items: 2<br>Reliability coeff.: 0.81 | During the current school year (2009–2010, including summer 2009), about how many total hours of PD did you receive so far in each of the following areas?<br>– Pedagogy: Strategies for teaching your subject area(s)<br>– Content: Indepth study of topics in your subject area(s) |
| On test and data<br>Items: 2<br>Reliability coeff.: 0.67 | During the current school year (2009–2010, including summer 2009), about how many total hours of PD did you receive so far in each of the following areas?<br>– Preparing students to take New York State assessments or Regents tests<br>– Analyzing and interpreting student achievement data |
| On nonacademic topics<br>Items: 2<br>Reliability coeff.: 0.81 | During the current school year (2009–2010, including summer 2009), about how many total hours of PD did you receive so far in each of the following areas?<br>– Classroom and behavior management<br>– Identifying and addressing student socioemotional issues |

**Table 8.4**
**SPBP Teacher Survey Scale Description: Collaboration**

| Construct | Description |
|---|---|
| **Collaboration scale:** | Never or not yet this year = 1<br>Once or twice a year = 2<br>Every few months = 3<br>Once or twice a month = 4<br>Once or twice a week = 5<br>Almost daily = 6 |
| Collaboration with other teachers<br>Items: 5<br>Reliability coeff.: 0.78 | During the current school year (2009–2010), how frequently have you engaged in the following activities?<br>– Analyzed student achievement data or student work with other teachers at my school<br>– Met with other teachers at my school to discuss instructional planning<br>– Met with other teachers at my school to discuss the academic needs of individual students<br>– Observed lessons taught by another teacher at my school<br>– Had my lesson observed by another teacher at my school |
| Collaboration with administrators<br>Items: 3<br>Reliability coeff.: 0.78 | During the current school year (2009–2010), how frequently have you engaged in the following activities?<br>– Analyzed student achievement data or work with administrators at my school<br>– Met with administrators at my school to discuss the academic needs of individual students<br>– Had an administrator observe my instruction (not for formal evaluation purposes) |
| Collaboration with whole staff<br>Items: 2<br>Reliability coeff.: 0.49 | During the current school year (2009–2010), how frequently have you engaged in the following activities?<br>– Discussed the social or behavioral needs of individual students with staff, including teachers, counselors, psychologists, etc.<br>– Met with whole school staff to discuss school goals and improvement strategies |

**Table 8.5**
**SPBP Teacher Survey Scale Description: Improvement Efforts**

| Construct | Description |
|---|---|
| **Improvement effort scale:** | Not at all = 1<br>To a small extent = 2<br>To a moderate extent = 3<br>To a great extent=4 |
| Collaborative planning<br>Items: 2<br>Reliability coeff.: 0.76 | To what extent have you undertaken the following to help your school achieve a high Progress Report score/grade this year?<br>– Worked together with other teachers to plan and implement what we teach<br>– Worked together with administrators to plan and implement what we teach |
| Attention to subgroups<br>Items: 2<br>Reliability coeff.: 0.64 | To what extent have you undertaken the following to help your school achieve a high Progress Report score/grade this year?<br>– Focused on the academic performance of students on the cusp of state test levels (e.g., between levels 2 and 3)<br>– Focused on the academic performance of high-needs students (e.g., ELL, SpEd, low performers) |
| Test preparation<br>Items: 3<br>Reliability coeff.: 0.74 | To what extent have you undertaken the following to help your school achieve a high Progress Report score/grade this year?<br>– Spent time in class teaching general test-taking strategies (e.g., time management, eliminating wrong multiple-choice options)<br>– Relied on particular styles or formats of problems in state tests in my own classroom assessments<br>– Focused on standards known to be assessed on state tests |

**Table 8.6**
**SPBP Teacher Survey Scale Description: Collegiality**

| Construct | Description |
|---|---|
| **Collegiality scale:** | Strongly disagree = 2<br>Disagree = 2<br>Agree = 3<br>Strongly agree = 4 |
| Collegiality<br>Items: 5<br>Reliability coeff.: 0.82 | Think about staff in your school this year, including teachers, other UFT members, and administrators. Please indicate the extent to which you agree or disagree about staff in your school —<br>– Feel responsible to help each other do their best<br>– Do not regularly share materials or ideas with one another<br>– Really trust each other<br>– Seem more competitive than cooperative<br>– Can be counted on to help out anywhere or anytime, even though it may not be part of their official assignment |

**Table 8.7**
**SPBP Teacher Survey Scale Description: Instructional Leadership**

| Construct | Description |
|---|---|
| **Instructional leadership scale:** | Strongly disagree = 2<br>Disagree = 2<br>Agree = 3<br>Strongly agree = 4 |
| Instructional leadership<br>Items: 6<br>Reliability coeff.: 0.92 | Think about the leadership that the principal at your school has provided this school year. To what extent do you agree or disagree with each of the following statements about your principal's leadership? The principal at my school —<br>– Sets high standards for teaching<br>– Carefully tracks student academic progress<br>– Ensures that teachers have sufficient time for professional development<br>– Actively monitors the quality of instruction in our school<br>– Works directly with teachers who are struggling to improve their instruction<br>– Communicates a clear vision for our school |

we adjusted for multiple comparisons.[2] This indicates that, overall, the program was not associated with differences in teachers' attitudes, perceptions or behaviors, with one possible exception: attitudes about whether nonteaching staff should receive a smaller share of a financial bonus than teachers.

## Findings

Overall, we did not observe statistically significant group differences on the survey measures, indicating that the assignment to the SPBP treatment or control group was not associated with differences in attitudes, perceptions, or behaviors.[3] In the next sections, we will review the results by key domain.

### Instructional Practice

As Tables 8.8 and 8.9 illustrate, we observed no statistically significant differences between SPBP and control-group teachers on most scales and individual items measuring instructional practices. Overall, both groups of teachers reported devoting similar amounts of time to test-score-driven instruction and student-directed instruction, to conferring with other teachers, and to introducing content through presentation or direct instruction. Two exceptions were that SPBP teachers reported devoting slightly less time to standards-aligned instruction than did control-group teachers—a finding that runs counter to the hypotheses that incentives might increase teachers' emphasis on certain types of instruction known to support student performance on state standards-based tests. In addition, a smaller percentage of the SPBP teachers than the control-group teachers reported that they used peer tutoring daily during instruction. However, neither of these estimates remained significant when, using our controls for multiple compari-

[2] We used the Benjamini and Hochberg, 1985, method to control for false significant rate at 5 percent across all 37 tests.

[3] For teacher outcomes, we also tested for interactions between school size and SPBP. We did not find significant interactions between them.

**Table 8.8**
**Comparison of Instructional Practice Scale Scores of Teachers in SPBP Treatment and Control Schools**

| Construct | N | Scale Mean | Coefficient | P–value |
|---|---|---|---|---|
| Test-score-driven instruction | | | | |
| Treatment | 814 | 3.56 | –0.02 | 0.72 |
| Control | 536 | 3.57 | | |
| Student-directed instruction | | | | |
| Treatment | 910 | 3.75 | –0.05 | 0.22 |
| Control | 601 | 3.79 | | |
| Standard-aligned instruction | | | | |
| Treatment | 910 | 4.40 | –0.09 | 0.01 |
| Control | 601 | 4.50 | | |

**Table 8.9**
**Percentages of SPBP Treatment and Control-Group Teachers Reporting Time Devoted to Instructional Strategies**

| Instructional Strategies | Group | N | Once or twice a month or week (%) | Almost daily (%) | P–value |
|---|---|---|---|---|---|
| Have students help other students learn class content (e.g., peer tutoring) | Treatment | 909 | 42 | 53 | 0.03 |
| | Control | 601 | 34 | 60 | |
| Confer with another teacher about alternative ways to present specific topics or lessons | Treatment | 910 | 60 | 30 | 0.24 |
| | Control | 601 | 64 | 26 | |
| Introduce content to the whole class through formal presentations or direct instruction | Treatment | 909 | 28 | 70 | 0.14 |
| | Control | 601 | 25 | 74 | |

sons, we accounted for the potential of one or more of the 37 tests to find spurious results. This suggests that the significant treatment effect we observed on these two outcome measures could represent spurious associations between treatment groups and outcomes, which is supported by the fact that no pattern of effects emerged.

### Teacher Effort, Collaboration, and Professional Development

On most scales and individual items measuring teacher effort, collaboration, and professional development, we observed no statistically significant differences between SPBP and control-group teachers. Overall, both groups of teachers reported spending a similar number of hours a week on school work outside the regular school day (mean for SPBP teachers was 12.80 hours; mean for control-group teachers was 13.17 hours). There were also no significant differences in the total number of hours spent in professional development (mean for SPBP teachers was 48.48 hours; mean for control-group teachers was 49.95 hours) and in the hours of professional development related to instruction, state tests and data, and nonacademic support for students (Table 8.10).[4]

Counter to the hypothesis that SPBP participation would increase staff collaboration, we observed no significant differences between groups in the frequency with which teachers reported collaborating with other teachers, with administrators, or with the whole school staff on instruction, school improvement, and student support (Table 8.11).

Finally, there were no significant differences in the efforts teachers reported to have undertaken to help schools achieve high Progress Report scores (Tables 8.12 and 8.13). Teachers in SPBP schools were just as likely as teachers in control schools to report undertaking collaborative planning, paying attention to particular groups of students, and engaging in test-preparation activities to help their schools

[4] As noted in Table 8.2, these items were measured on a 5-point scale: None = 1, 1 to 5 hours = 2, 6 to 20 hours = 3, 21 to 40 hours = 4, More than 40 hours = 5. Therefore, a mean score of 2.48 for treatment schools on time spent in instructionally related professional development means that teachers on average reported receiving between 1 to 5 hours and 6 to 20 hours of this type of professional development.

**Table 8.10**
**Comparison of Teacher Time Devoted to Non–School-Hour School Work and Professional Development in SPBP Treatment and Control Schools**

| Construct | Group | N | Mean | Coefficient | P–value |
|---|---|---|---|---|---|
| Time spent on work outside regular school hours (hours per week) | Treatment | 899 | 12.80 | –0.34 | 0.46 |
| | Control | 596 | 13.17 | | |
| Overall hours on professional development this school year | Treatment | 987 | 48.48 | –1.78 | 0.83 |
| | Control | 595 | 49.95 | | |
| Time spent in professional development: instruction | Treatment | 894 | 2.48 | –0.07 | 0.22 |
| | Control | 596 | 2.54 | | |
| Time spent in professional development: test and data | Treatment | 894 | 2.31 | –0.01 | 0.92 |
| | Control | 596 | 2.32 | | |
| Time spent in professional development: nonacademic | Treatment | 894 | 1.77 | 0.00 | 0.94 |
| | Control | 592 | 1.77 | | |

**Table 8.11**
**Comparison of Teacher Collaboration in SPBP Treatment and Control Schools**

| Construct | Group | N | Scale Mean | Coefficient | P–value |
|---|---|---|---|---|---|
| Collaboration with | | | | | |
| Other teachers | Treatment | 891 | 3.43 | –0.04 | 0.47 |
| | Control | 592 | 3.47 | | |
| Administrators | Treatment | 891 | 3.02 | –0.03 | 0.59 |
| | Control | 592 | 3.05 | | |
| Whole staff | Treatment | 891 | 3.53 | 0.01 | 0.80 |
| | Control | 592 | 3.50 | | |

**Table 8.12**
**Comparison of Three Types of Teacher Efforts to Help Schools Achieve High Progress Report Scores in SPBP Treatment and Control Schools**

| Construct | Group | N | Scale Mean | Coefficient | P–value |
|---|---|---|---|---|---|
| Collaborative planning | Treatment | 874 | 3.18 | –0.10 | 0.05 |
| | Control | 586 | 3.27 | | |
| Attention to subgroups | Treatment | 874 | 3.26 | 0.01 | 0.87 |
| | Control | 586 | 3.25 | | |
| Test preparation | Treatment | 874 | 3.11 | 0.03 | 0.55 |
| | Control | 586 | 3.10 | | |

**Table 8.13**
**Percentage of SPBP Treatment and Control-Group Teachers Reporting Undertaking Efforts a Moderate or Great Extent to Help School Achieve High Progress Report Scores**

| Item | Group | N | Percentage | P–value |
|---|---|---|---|---|
| Sought out PD opportunities to improve my practice | Treatment | 874 | 74 | 0.34 |
| | Control | 585 | 74 | |
| Devoted effort to improving student attendance | Treatment | 873 | 81 | 0.24 |
| | Control | 586 | 85 | |
| Offered assistance outside of school to help students who were struggling | Treatment | 873 | 57 | 0.31 |
| | Control | 585 | 54 | |
| Worked jointly with students to set and monitor individual goals | Treatment | 874 | 79 | 0.21 |
| | Control | 583 | 82 | |
| Provided mentoring and other support to fellow teachers | Treatment | 871 | 65 | 0.53 |
| | Control | 583 | 62 | |
| Communicated with parents about the importance of completing survey | Treatment | 872 | 64 | 0.51 |
| | Control | 582 | 64 | |

improve.[5] There were also no differences in the extent to which teachers reported focusing on other areas, such as student attendance and communicating the importance of parent survey responses—two measures contributing to the Progress Report scores and thus to bonuses. As described in Chapter Six, some case-study SPBP staff had reported increased attention to these two areas, but these teacher survey results indicate that such efforts were not unique to schools eligible for the bonus and may have been more universal to all schools subject to Progress Report accountability measures.

### Attitudes

As described earlier, understanding of and buy-in to the measures used to determine the schoolwide bonus may have mediated the effects of the financial incentive. Interestingly, there were no significant differences between SPBP and control-group teachers in the reported level of understanding of the Progress Reports and the factors contributing to their schools' scores (Table 8.14). In fact, both reported high levels: 85 percent of the treatment group and 86 percent for the control-group teachers agreed or strongly agreed that they had a strong understanding. Further, both sets of teachers expressed similar levels of concern about basing financial bonuses on test scores: 82 percent of the treat-

**Table 8.14**
**Percentage of SPBP and Control-Group Teachers Agreeing or Strongly Agreeing with Statements About Progress Reports and Rewards**

| Item | Group | N | Percentage | P–value |
|---|---|---|---|---|
| I have a strong understanding of our school's Progress Report and what factors contribute to our school's score this year (2009–2010) | Treatment | 865 | 85 | 0.60 |
| | Control | 591 | 86 | |
| Rewarding teachers based on test score gains is problematic because it is hard to relate gains in student achievement to the work done by teachers | Treatment | 841 | 82 | 0.21 |
| | Control | 586 | 85 | |

[5] Although the differences between groups on collaborative planning had a p-value of 0.05, the results were not statistically significant after adjusting for multiple comparisons.

ment group and 85 percent of the control-group teachers agreed or strongly agreed that rewarding teachers based on test score gains is problematic because it is hard to relate gains in student achievement to the work done by teachers.

When asked about how to distribute a schoolwide bonus, teachers in the SPBP and control schools reported similar preferences for distributing equal shares to all staff (Table 8.15).[6] However, SPBP teachers were significantly more likely to agree that nonteaching staff should receive a smaller share of the bonus than teachers, suggesting a slightly less egalitarian position among SPBP teachers than among control-group teachers. This difference between the SPBP and control schools remained statistically significant after we adjusted for multiple comparisons.

### Teacher Mobility and Retention

As Table 8.16 illustrates, there were no significant differences between SPBP and control groups in the percentage of teachers reporting an intention to stay at the same school for the subsequent school year.

**Table 8.15**
**Percentage of SPBP and Control-Group Teachers Agreeing or Strongly Agreeing with Statements About Distributing Financial Bonuses**

| Item | Group | N | Percentage | P–value |
|---|---|---|---|---|
| All building personnel should receive an equal share of the bonus | Treatment | 844 | 65 | 0.32 |
| | Control | 581 | 70 | |
| Nonteaching staff should receive a smaller share of the bonus than teachers | Treatment | 803 | 61 | 0.00 |
| | Control | 582 | 56 | |
| Nonteaching staff should not be eligible for a share of the bonus | Treatment | 809 | 25 | 0.15 |
| | Control | 582 | 24 | |

[6] Control-group teachers were asked about a hypothetical situation, and statements were prefaced with: "In a system of schoolwide bonus awards based on our school's Progress Report score . . . ." Treatment teachers were simply asked to respond to questions about the SPBP, and items were identical to those listed in Table 8.11. Although the statements for both groups were not equivalent, the comparison of opinions are still valuable.

**Table 8.16**
**Percentage of SPBP and Control-Group Teachers Agreeing or Strongly Agreeing with Statements About Teacher Mobility**

| Item | Group | N | % Yes | P–value |
|---|---|---|---|---|
| Do you plan to return to this school next year? | Treatment | 737 | 94 | 0.13 |
| | Control | 519 | 96 | |
| Among teachers responding No above — | | | | |
| I applied for a teaching position at a different school within the district | Treatment | 188 | 16 | 0.86 |
| | Control | 92 | 16 | |
| I applied for a teaching position at a different school in another district | Treatment | 188 | 24 | 0.45 |
| | Control | 91 | 30 | |
| I applied for a nonteaching job (including noneducation job) | Treatment | 184 | 8 | 0.38 |
| | Control | 90 | 7 | |
| I received a formal job offer to teach in another school | Treatment | 185 | 7 | 0.79 |
| | Control | 91 | 9 | |
| I received a formal job offer for a non–teaching job (including non–education job) | Treatment | 183 | 4 | 0.73 |
| | Control | 91 | 3 | |
| I am retiring or may be retiring | Treatment | 183 | 10 | 0.80 |
| | Control | 93 | 12 | |

Similarly, among teachers reporting a possibility of not staying at their current school, SPBP teachers were just as likely as control-group teachers to report submitting applications for teaching positions at a different school in either the same district or another district or receiving formal job offers to work elsewhere.

### School Environment

Despite the frequently expressed concern among educators and policymakers that financial incentives could damage the collegial environment in schools, the bonus-eligible teachers in this study tended to report similar school climate conditions as their control-group colleagues (Table 8.17). Similarly, there were no significant differences in teacher reports of instructional leadership of their principal.

**Table 8.17**
**Comparison of School Environment Indicators (Collegiality and Instructional Leadership) in SPBP Treatment and Control Schools**

| Construct | Group | N | Scale Mean | Coefficient | P–value |
|---|---|---|---|---|---|
| Collegiality | Treatment | 890 | 2.89 | –0.03 | 0.48 |
| | Control | 593 | 2.92 | | |
| Instructional leadership | Treatment | 889 | 2.95 | –0.02 | 0.62 |
| | Control | 591 | 2.96 | | |

## Summary and Discussion

Overall, we found that SPBP did not affect self-reported teacher attitudes, perceptions, and behaviors. After controlling for multiple comparisons across all 37 measures, we found just one statistically significant difference between teachers in treatment and control schools. Teachers in the SPBP schools reported similar instructional practices, levels of effort, staff collaboration, types and amounts of professional development, attitudes about rewards, intentions to stay at their current schools, and school climate conditions as their control-group colleagues. These findings indicate that the assignment to the SPBP treatment or control group was not associated with differences in reported attitudes, perceptions, or behaviors.

CHAPTER NINE

# Conclusions and Implications

This study was designed to evaluate the implementation and effects of the New York City SPBP. Drawing on case studies, interviews, surveys, administrative data, and documents, along with an analysis of Progress Report and student achievement data, we examined the history and roll-out of the program, the experiences of teachers and CCs, the bonus distribution plans developed, overall attitudes about the program, effects on schools and individuals, and the effects on student achievement. In this final chapter, we summarize our key findings about the program, present a set of recommendations for NYCDOE and UFT leaders, and discuss implications for pay-for-performance policy more generally.

## Key Findings

### The Study Found SPBP Did Not Improve Student Achievement

**Overall, SPBP Did Not Improve Student Achievement at Any Grade Level.** Analyses of student achievement on the state's accountability tests found small negative differences between the average achievement of students attending SPBP treatment and control elementary, middle, or K–8 schools in years 1 through 3. These differences, however, were statistically significant only for mathematics in year 3, and none were significant when we controlled for multiple testing. These results were robust under various analytic approaches. Similarly, there were no statistically significant overall effects on Regents Exam scores

for high school students in years 1 and 2. The team tested for but did not find differential program effects by school size and found no relationship between student achievement and the CC distribution plans for bonuses among staff.

**SPBP Did Not Affect School Progress Report Scores.** Across all years and all the component scores for the Progress Reports (environment, performance, progress, and additional credit), the researchers found no statistically significant differences between scores of SPBP treatment and control schools and between schools that participated in SPBP each year (regardless of random assignment) and other eligible schools. This lack of effects held true for elementary, middle, and high schools.

### The Implementation Had Mixed Success in Creating the Optimal Environment for SPBP

SPBP implementation depended on communication about the program to participating schools and their staffs, the CC determination of award distributions, the determination of award winners, and the subsequent payout of those awards. Past research on pay-for-performance programs and expectancy theory (Vroom, 1964) suggests that, to achieve the desired program results, these activities must follow a reasonable timeline and lead to a high degree of the following key attributes:

- *understanding* of the program, as shown by knowledge of criteria by which incentives are awarded and the amount of money at stake
- *expectancy*, as demonstrated by educators' beliefs that they are capable of doing things that will enable them to achieve the targets
- *valence*, a concept that refers to the belief that incentives are sufficiently valuable or substantial to inspire responses predicted by the theory of action
- *buy-in*, or acceptance of the program and its criteria
- *perceived fairness.*

The findings below describe the rollout, communication, and the extent to which the implementation achieved the key attributes necessary for success.

**Basic Procedures Were Enacted as Planned.** NYCDOE and UFT generally implemented SPBP as intended. At the start of each year, schools voted to participate, and later in the year, formed their CCs and submitted bonus distribution plans. Schools had the freedom to compensate staff with bonuses as they desired, without interference from either NYCDOE or UFT. Each year, NYCDOE awarded bonuses according to the program guidelines. In the first two program years, district leaders announced the bonuses within the first few months of the next school year. However, in year 3, announcements were delayed until nearly midway through the subsequent school year.

**Communications Followed the SPBP Design, with Early Problems Being Corrected Over Time.** Both NYCDOE and UFT adhered to the communication plan for SPBP, which called for sharing joint DOE-UFT written materials about the program with school staff. According to survey respondents, UFT representatives served as the main conduits for information about SPBP to UFT CC members and teachers. CC respondents reported early misunderstandings but that communications improved over time. Nevertheless, some interviews and teacher survey responses suggested that some misunderstanding of program components remained into year 3, when UFT and DOE reduced the emphasis on communicating about the program.

**The CC Process Was Implemented Fairly and Smoothly, but Some Schools Had Difficulty with the Decisionmaking Process.** Each school formed a four-person CC and generally followed SPBP guidelines regarding membership and procedures. Most CC members reported that the decisionmaking process was fair, collegial, and inclusive, and that achieving consensus was easy; however, some survey and case-study respondents expressed concerns about the process. For example, 44 percent of teachers disagreed or strongly disagreed that teachers' preferences were taken into account when developing the distribution plan. Some CC members whom we interviewed also questioned whether the requirement that UFT-represented staff make up only one-half of the members truly guaranteed an even playing field and spoke

about power differentials that played out between administrators and UFT-represented staff in committee deliberations.

**The Majority of CCs Developed Nearly Egalitarian Award Distribution Plans, Reflecting Strong Preferences Among CC Members That Staff Members Share Bonuses Equally.** Although administrators were significantly more inclined than their UFT counterparts to favor differentiating bonuses so that different staff members would be eligible for different amounts, there was a strong preference overall for egalitarian plans among CC members. Further, almost two-thirds of teachers indicated a preference for distributing equal shares of the bonus to all school staff (yet almost the same proportion of teachers also reported that nonteachers should receive a smaller share than teachers). Not surprisingly, the majority of committees developed equal-share distribution plans in both years. The most common individual bonus amount awarded to staff in the plans was $3,000 in all three years. Most staff within a school received the same award amount. In fact, in 2010, 82 percent of staff members were slated to receive the most common or modal award for their school, and even greater equality existed in the early years of the program. Inconsistent with the notion that larger schools might use differentiation in bonus payments to offset "free riders" (i.e., staff who shouldered less than their fair share of the work but still collected a full bonus), award equality increased with school size.

About 31 percent of schools reported using individual performance as at least one of the factors for determining awards. The remaining schools either did not differentiate or reported using only factors related to time or job title but not individual performance. Unequal disbursement at times led to resentment within the schools, and some schools with highly differentiated allocation plans one year adopted much more egalitarian plans the subsequent year.

**The Few Schools That Determined Bonus Shares by Individual Performance Tended to Have More Differentiation Among Award Amounts, but Did Not Differ from Other Schools in Student Achievement.** Sixty schools reported including individual performance measures, such as staff absences, unsatisfactory staff evaluation ratings ("U-ratings"), or other performance measures as a factor (among many

others) for determining individual staff awards. Compared to other schools that did not report relying on individual performance measures in developing their distribution plans, these schools were more likely to award staff less than the modal awards and more likely to award certain staff members no money. According to the theory of action for SPBP that some leaders espoused, greater disparity in awards would help to incentivize performance, have a motivational effect on individual behavior, and increase student achievement. However, even the schools that determined awards by individual performance generally remained cautious about deviating from egalitarian awards and slated 74 percent of staff, on average, for the modal award amount. Moreover, students in these schools—whether elementary, middle, or K–8—did not have higher achievement than those in other SPBP schools.

**A Minority of CC Members and a Few Case-Study Respondents Reported Problems with the Distribution of Bonuses.** The distribution process was reported to run fairly smoothly in year 3, and only a minority of CC members reported problems, such as staff being left off the list of recipients or being upset with a perceived unfair distribution. None of these reported problems reached the level of official appeal in any year of the program. Participants in several case-study schools reported dissatisfaction among staff when decisions and criteria used to differentiate the bonus were viewed as opaque and subjective or when the schools did not communicate the final distribution plans to staff members.

**The Implementation of SPBP Had Mixed Results in Creating the Conditions that Foster Success.** We noted the following:

- *Understanding and Awareness.* Staff members reported being aware of the program and generally supportive of it. Most educators surveyed and interviewed greatly appreciated the financial reward and the recognition for their efforts. However, there were persistent misunderstandings about the Progress Reports and other program elements. According to surveys, more than one-third of teachers did not understand key aspects of the program, including the target their schools needed to reach, the amounts of money their schools would receive if they met their targets, the

source of funding, and how committees decided on distribution plans. In case-study schools, many individuals conveyed misperceptions about various aspects of the programs, and many called for better communication.

- *Expectancy.* In year 3, staff showed some accuracy in their assessment of the efforts required to earn a bonus but, overall, seemed to overestimate the likelihood that their schools would receive bonuses. For instance, although many CC members acknowledged needing to improve performance to win a bonus in 2009–2010, a large majority felt certain their school would receive one.
- *Valence.* The majority of teachers and CC members expressed a strong desire to win bonuses and found the financial bonuses motivating, but many winners reported that, after taxes, the bonus seemed insignificant. In fact, almost one-half of the teachers responding to the survey and some staff members in case-study schools indicated that the bonus was not large enough to motivate extra effort. Further, many case-study respondents reported viewing the bonus as a reward for their usual efforts, not as an incentive for changing their behavior.
- *Buy-in.* Buy-in for the program performance measure was limited: The majority of teachers and CC members felt the criteria relied too heavily on test scores.
- *Fairness.* More than one-half of teachers and CC members felt the program was fair to participating schools, and CC members did not report dissatisfaction about unfair distributions. However, some UFT-represented staff expressed dissatisfaction with the composition of the CC committee; more than one-half of teachers and UFT committee members wanted UFT to have greater than 50-percent representation on the committee. Some staff also felt the Progress Report targets were too high.
- *Timelines.* Three-fourths or more of teachers and CC members suggested that they should have been informed of the distribution plans at the start of the year.

### SPBP Did Not Produce the Intended Effects on Teachers' Reported Attitudes and Behaviors

**SPBP Did Not Affect Teacher Reported Attitudes, Perceptions, and Behaviors.** The survey found almost no differences between the practices and opinions of teachers in the SPBP group and those of the control group. On all measures related to instructional practices, effort, participation in professional development, mobility, and attitudes, the responses from the two groups of teachers were very similar, and there were no statistically significant differences among them.

**The Lack of Results Might Be Due to the Limited Motivational Power of the Bonus.** The theory of action behind SPBP was that the potential for a bonus would motivate change, but this study did not find such change. The vast majority of teachers and CC members who received bonuses said that winning the bonus was a nice acknowledgment of their hard work but that it did not influence their performance. In addition, only about one-third of CC members and 16 percent of teachers reported that not receiving a bonus energized them to improve their practices the subsequent year, and only a very small proportion of both groups actually reported that not receiving the bonus reduced their motivation.

The limited motivational effect of the bonus might have resulted from school staff viewing the award as a reward rather than an incentive: It made them feel appreciated for their hard work, but they claimed that they would have undertaken this work with or without the bonus. The size of the award might also have been a factor. Some stakeholders believed that differentiation among awards could enhance the motivational power of the bonuses. As discussed previously, differentiation especially related to performance was very limited, and it is possible that this too contributed to the limited incentivizing effects of the bonus. Finally, other accountability pressures and intrinsic motivations were often perceived to be more salient than the bonus. For example, although 64 percent of teacher survey respondents said the possibility of earning a financial bonus motivated them moderately or to great extent to work hard as a teacher, an even greater percentage reported the same motivational value from the possibility of receiving a high Progress Report grade (without the financial incentive attached)

(75 percent) and achieving or exceeding their school's AYP target (77 percent).

**The Lack of Effect Was Also Echoed in Responses About Staff Members' Own Performance.** The majority of teachers, CC members, and case-study respondents said that SPBP did not affect their own job performance or motivation.

**It Is Less Clear Whether the Lack of Effects on Collaboration or Morale Contributed to the Lack of Intermediate Outcomes.** SPBP's theory of action hypothesized that the chance to earn a bonus on the basis of school performance could enhance collaboration and improve outcomes. It also suggested that winning bonuses could boost morale and improve outcomes. However, the data on collaboration and morale were mixed. More than 40 percent of participating teachers and CC members reported that teachers' willingness to collaborate and work together, teachers' focus on student learning, and staff morale changed for the better as a result of participating in SPBP. The perception of increased willingness to collaborate, however, was not reflected in the comparison with schools not in the program. The reported levels of staff collaboration differed very little between SPBP and control schools, and the differences that existed tended to show greater collaboration in the control schools (but the differences were not statistically significant).

## Discussion

The findings from the Progress Report and student achievement analyses suggest that in the first three years of implementation SPBP did not achieve its ultimate goal of increased student performance. There are several plausible explanations for these results.

First, one could argue that the program was too new to expect to see effects. As noted, the program was first introduced midyear in 2007–2008, three months before state tests were administered, and many staff members may not have been aware of their school's participation or the incentive until the end of the year or when they received the actual bonus payment the next fall. In year 2, assuming that staff

were not sure their school was going to participate at the start of the year when a vote was required, the January 2009 administration of the ELA state test may have also given school staff only a limited time to operate within the bonus program prior to testing. Thus, in the first two years, it is reasonable to ask when teachers could have even begun to be motivated by the bonus, and how many months after that point would they have had to actually influence the performance measures? Is it realistic to expect that school staff could have improved instructional quality and student outcomes in that limited time? The lack of positive results found in year 3, however, is not as easily accounted for by this explanation, since more staff members should have been aware of the program and since there was at least a reasonable window of opportunity for staff to experience motivational effects. One could still argue, nevertheless, that it may take longer than three years for changes in teachers' practices to result in higher test scores. Yet, the lack of teacher effects described in Chapter Eight provides no support for this argument. Our results suggest that, after three years, the program was having no significant effect on teacher behavior.

Second, the findings from the teacher and CC surveys and case-study interviews suggest that several factors important for pay-for-performance programs might not have been in place in all participating schools. The absence of these factors might have weakened the motivational effects of the bonus. Circumstances that could have weakened SPBP's effects include misunderstandings about the program, uncertainty that teachers could sufficiently change practices or act in ways to achieve targets, lack of buy-in for the bonus criteria, concerns about timing, and questions about procedural fairness, among others. Without a high degree of understanding, expectancy, valence, buy-in, perceived fairness, and a reasonable timeline, it is not surprising that SPBP was not a strong motivational and reform tool in all the schools.

The lack of observed results could also be due to the low motivational value of the Progress Reports and the accountability incentives all schools faced. Assessing the true motivational value of the bonus is a difficult task. Our data seem to yield mixed findings on this topic. While many teachers and other staff reported that the possibility of receiving a bonus did not affect their practices, many none-

theless expressed a desire to earn the bonus and indicated that they were taking it seriously. Nevertheless, many acknowledged that other accountability pressures and incentives (e.g., receiving a high Progress Report grade, achieving AYP targets) held the same motivational value as the possibility of receiving a financial bonus.

Even though winning the bonus was desirable, that does not mean it motivated teachers to change behaviors. Teachers reported that the bonus motivated them to work hard but maybe not harder than they would have in its absence. The desire to receive a good Progress Report grade was even more broadly a strong motivator and was an incentive for all staff in all schools both SPBP and control. Hence, while the bonus might have been another factor motivating SPBP staff to work hard, they would have had similar motivation without it because of the high level of accountability pressure on all schools and their staffs. Consequently, SPBP might not be expected to change behavior or to influence student outcomes.

Several other programs (e.g., Teachers of Tomorrow Program, Opportunity New York City-Family Rewards program) may have also been in place during the study period that rewarded teachers and students for improved performance. These programs included teachers and students in both SPBP and control schools. Like the general accountability pressures, these other programs incentivized all participants to change behavior and could have limited the ability of SPBP to yield any additional behavioral changes over and above those already being made.

Some may also argue that, because of the lack of differentiated performance plans in most schools, the true motivational potential of SPBP was not adequately realized. Some NYCDOE leaders clearly intended schools to seize on the autonomy of the CC to distribute funds in ways that provided additional, individual-based performance incentives. Over the three years, very few schools opted for differentiated distribution plans, and many differentiated in ways not based on individual performance but instead on such factors as time spent at the school or job titles. We found that student achievement was not greater in schools that differentiated in part on the basis of individual performance, even though these schools had significantly greater disparity in

their awards than other schools. However, the sample of schools was very small, giving us little statistical precision to estimate the effects of such distribution strategies.

Another possible explanation for the lack of positive student achievement effects is a flaw in or "incompleteness" regarding the underlying theory of action for SPBP and school-based pay-for-performance programs more generally. As some have argued in the past, motivation alone may not improve schools (Elmore, 2004; McLaughlin, 1987). Even if the bonus inspired staff to improve their practices or work together, staff may not have the capacity or resources (e.g., school leadership, social capital, expertise, instructional materials, time) to bring about improvement. As one researcher has argued in the past about other school-based pay-for-performance programs, "even schools that may be inspired by the school-based incentive system may be stuck or stymied. They could be more willing but still not able to improve" (Malen, 1999, p. 390).

Finally, it is also worth noting the contradictions between staff perceptions of program effects reported in Chapter Six and the more direct effects reported in Chapters Seven and Eight. There are several possible explanations for this misalignment. First, it is possible that survey respondents were reporting socially desirable responses and thus inflating the perceived effects of the program. Second, it may be that staff in SPBP schools attributed changes in motivation and outcomes to SPBP participation but that other factors led to similar results in the control schools. For example, if SPBP had not been in place, teachers in SPBP schools might have been equally motivated by the desire for a good Progress Report grade. Third, the effects that teachers report may have occurred but been too small to change outcomes or been identified through our analyses. Fourth, it is possible that at least some of the teacher reports are accurate but that the factors that changed were not related to student outcomes and, consequently, did not translate into program effects on student achievement. Even though the reports were not the best measure of the causal effects of the program on certain behaviors and student outcomes, they did indicate a level of positive response to the program and to pay-for-performance, since staff not

only reported positive perceptions of the program but were also willing to attribute positive effects to it.

## Recommendations

Overall, evidence from this study did not support continuing SPBP. The program did not have the desired effects on student achievement or the necessary intermediate teacher outcomes. There was no consistent evidence that the program motivated educators to change their practices or that continuing the program would improve outcomes. Moreover, the program was costly, and its costs fluctuated dramatically across years.

As we discussed earlier, some implementation problems may have weakened the effects of the bonus and the program overall, and NYCDOE and UFT leaders could consider the following actions to address those if they wish to consider a similar program again:

### Improve Communication

NYCDOE and UFT leaders may want to consider disseminating information early in and regularly throughout the school year to ensure that all staff know how bonuses are calculated and on what basis, the dollar amounts at stake, and the general process for how the program operates. Other areas to clarify are allowable options for distributing funds and who is eligible (e.g., who can be added to the TO, staff pooling of funds for non-UFT staff). Further clarification about the role that personnel evaluation data can play in the CC deliberations and the potential privacy issues that might arise with such use would also be warranted. Some respondents requested that UFT leaders from headquarters come to schools to discuss the program with staff.

Leaders may also want to consider centralizing communication with CC members and participating schools by placing one person in charge of disseminating the same information to all participants. Regular communications about the program throughout the year might also improve understanding and buy-in. NYCDOE and UFT might also consider requiring schools to form their committees at the start

of the year and submit these names and email addresses. This would permit communication with CC members much earlier in the school year. Both these changes might address the differences in the reported receipt and perceived adequacy of information received by UFT staff and administrators serving on CCs. Another option would be to maintain a website where teachers interested in learning more about SPBP can access information and any materials that were distributed to school building designees.

School staff may also benefit from information about the committee process, for example, the importance of electing representatives to serve who understand staff preferences, are skilled at deliberating, and know how to "hold their own" in a group of individuals who may possess different viewpoints or levels of power. Leaders may also want to consider the potential value of improving communication about the program to parents and external constituencies, who currently are not the target of communication efforts within schools. It is possible that by knowing more about the program these stakeholders could not only provide additional "productive" pressure and incentive to improve, but perhaps contribute resources and help build capacity to help schools improve and attain their targets.

#### Adjust the Program Timeline to Ensure That All Staff Know at the Start of the Year the Distribution Plan and Dollar Amounts They Can Potentially Earn

If schools are required to develop their plans at the beginning of the academic year, staff would have a better understanding of what is at stake and might think about bonuses more as an incentive than simply a reward. One drawback, of course, is that if schools are required to submit each staff members' name and dollar amount at the start of the year, committees would not be able to use current school year performance data (e.g., periodic assessment results, personnel evaluations) as the basis for differentiating bonus awards. However, given that so few committees seemed to use performance data to award bonus shares in all three years, this might not be a compelling reason to maintain the existing timeline. Nevertheless, one way to address this concern is to require start-of-year plans to include the general criteria or basis

for awarding bonuses (e.g., individuals will receive more if they obtain higher teacher evaluation ratings, produce higher interim assessment results, serve the full year at the school), but leave the final decisions about allocating specific dollar amounts until the end of the year. It might also help to clearly communicate at the start of the year that bonuses are taxed to avoid the disappointment that many appeared to experience in the past.

### Ensure Transparency of Distribution Criteria and Plans

Related to the first two recommendations, NYCDOE and UFT leaders should consider requiring all committees to disseminate distribution plans and publicize the rationale for their decisions to all staff. These efforts could potentially prevent some of the conflict that arose in case-study schools that differentiated payouts without explaining to staff the criteria for these decisions. Required dissemination of distribution plans would also more clearly signal to staff what is being rewarded and, where relevant, what staff could do to potentially earn a greater share of the bonus pool.

### Provide Technical Assistance to CC Members to Facilitate Democratic Decisionmaking

To ensure that decisionmaking is truly democratic, it might help if NYCDOE and UFT leaders provided guidance about rules and procedures to follow when deliberating about bonus distribution. While schools that decide on equal-share plans may not need this support, others that debate differentiated options could benefit from information on how to weigh options, conduct open debate, and ensure that all members have an equal say in the final decisions. Such support might include information on how to establish norms and rules of engagement, with specific guidelines on how to ensure reasoned and democratic procedures and methods for deliberating and achieving consensus. Such groups as the National Coalition for Dialogue and Deliberation or organizations supporting Critical Friends Groups may have protocols and other tools that help establish these norms, rules, and procedures.

### Final Thoughts on Recommendations

Of course, other recommendations are less feasible, given current budget constraints. For example, although increasing the average size of bonuses might strengthen the perceived value and motivational effects of the bonus, it is not practical during a period of budget cutting and fiscal constraint. Similarly, it is possible that adjusting Progress Report measures might also improve buy-in for some staff members. For example, reducing the weight given to survey results might remove any incentives that exist for "gaming" this aspect of the Progress Report score and prevent efforts to artificially inflate results. However, survey results provide one of the few input measures and, if removed, may heighten existing concerns that targets rely too heavily on test-score results. To this point, many educators appear eager to include other measures of student outcomes that are not test based. Again, making such adjustments would require significant resources that may not be available.

However, we have no evidence that implementing these adjustments would improve the effectiveness of the program significantly. Theory suggests that good communication, timeliness, and transparency are necessary for programs like SPBP to succeed, but it does not guarantee that the awards will have the desired motivational incentives or lead to changes in school staff or student outcomes. Furthermore, given the consistent failure of recent experiments with similar bonus programs to positively affect student outcomes, we suspect a more-comprehensive revision of the program and reevaluation of the theory of action is likely to be necessary to achieve the desired outcomes.

Finally, the sharp decrease in the percentage of schools receiving bonuses in year 3 (13 percent) compared to year 2 (84 percent) also suggests a broader lesson for policymakers generally. Although the study did not collect data on responses to the year 3 results, it is hard to imagine how such fluctuations could improve the already weak motivational effect of the program. Prior to implementing a performance-based bonus program, it would behoove leaders to ensure greater stability of the accountability measures on which the bonuses rely.

## Implications for Pay-for-Performance Policy

The results of this evaluation add to a growing body of research from the United States that finds no effects on student achievement of narrow pay-for-performance policies that focus only on financial incentives without such other features as targeted professional development or revised teacher evaluations. In recent years, a few studies, released well after SPBP started, have demonstrated that several school- and individual-based performance incentive policies had no significant effect on student test results (e.g., Glazerman and Seifullah, 2010; Springer et al., 2009a; Springer et al., 2010). The New York City SPBP provided one of the few examples of a school-based incentive policy and, like the individual programs, had few observable effects on students, staff, and schools. The implementation findings, nonetheless, provided insights into the theory of action underlying the policy and the mechanisms expected to yield positive student outcomes—such as improved motivation, enhanced morale, increased collaboration, and improved practices. The study also uncovered much about the process and potential challenges of assigning a small school committee the task of distributing a schoolwide bonus among staff.

Overall, these results yielded several implications relevant to the broader set of pay-for-performance policies that have gained considerable national attention and interest in recent years:

- *Conditions must foster strong motivation.* Our study findings support much of the theoretical literature and other empirical research suggesting that there may be a set of key conditions needed to bolster the motivational effect of financial incentives, including a high degree of understanding about the program (e.g., criteria for awarding incentives, the amount of money at stake), expectancy (i.e., belief that educators can sufficiently act in ways to achieve targets), valence (i.e., belief that incentives are sufficiently valuable or substantial to inspire responses predicted by the theory of action), buy-in (e.g., for the criteria by which incentives are awarded), perceived fairness, and a reasonable timeline. Several of these purported key system components were lacking in SPBP

and were identified by educators as limiting the ability of program to change their behaviors.

Some of these conditions, such as expectancy, understanding, and buy-in, are also closely tied to the broader accountability system. For example, some of the confusion and resistance educators expressed may relate to the Progress Reports alone, separate and apart from the bonus. Many did not like the overreliance on test results to measure success, and many did not understand the complicated calculations used to arrive at the final Progress Report scores. These concerns suggest the need to invest in broader communication around the accountability system and to consider the level of buy-in for and understanding of those measures. Staff may be unwilling to work toward a goal they do not support, even if they find bonuses valuable. It may be necessary for the system to engage staff in developing the performance metrics and be willing to revise them if they do not receive general support or retain any initial buy-in. Additional support also might be needed to help staff motivate to improve to achieve its goals.

Other related conditions that might mediate the motivational signal of incentives that were not systematically examined in this study are the capacity of educators to achieve the accountability targets and the organizational context in which educators operate. If educators feel that they do not have the resources, or that organizational or political barriers impeded their ability to act in ways that will help them achieve their targets, then incentives will be of little use. Such a scenario might require investments in capacity building and organizational structure, policies, and cultures that facilitate improvement.

- *It is important to identify the factors that truly affect motivation.* Motivation is a key to the theory of change in pay-for-performance programs. The failure of recent experiments to find effects (Springer et al., 2010a, b) has generally been attributed to the lack of motivational effects in such programs, which is consistent with research studies in other sectors (Marsden and Richardson, 1994; Milkovich and Wigdor, 1991; Pearce and Perry; 1983). However, the in-depth data from this evaluation present a somewhat con-

tradictory story. Teachers and other staff reported that the possibility of earning a bonus motivated them to work hard and was desirable, but they also reported not changing their behaviors in response to the program. It may be that a desirable award is not enough to actually change behavior. This may be particularly true in the context of high-stakes and high-profile accountability. More teachers reported that receiving good Progress Reports is a motivator than that the bonus motivated them. The Progress Report has no direct financial rewards, but it has high stakes in terms of public opinion and possible negative sanctions. All schools face these accountability pressures; thus, these pressures may make the value of the additional incentive of the bonus too low to matter, or the bonus may just substitute for other motivational factors.

It may also be that concerns about experiencing negative sanctions—again a factor faced by all schools—is a greater motivator of true change than a bonus with unknown expected value. The experience of the POINT experiment in Nashville (Springer at al, 2010) is consistent with this conjecture. In the context of high-stakes accountability—schools in the district were facing possible state takeover—student outcomes of both treatment and control teachers demonstrated notable improvements, but the additional effects of the bonuses on student achievement were limited to none. Finally, as noted earlier, the limited motivational effects of the bonus that teachers and other school staff reported might not be enough to explain why SPBP failed to change teacher behaviors. Such factors as lack of capacity to change or to identify alternative teaching practices might have prevented change, even if the limited motivation of the SPBP bonus induced teachers to try to improve.

- *Performance-based incentives may face challenges from the micropolitics of school-level implementation.* This evaluation highlighted the underlying political tensions inherent in implementing a bonus system. Although many major program elements were implemented smoothly across participating schools (e.g., the formation of CCs), some schools found it difficult to decide how to distribute

bonuses among staff (e.g., power differentials played out between administrators and UFT-represented staff during deliberations). The majority of CCs distributed bonuses nearly equally among all staff, and some unequal disbursements exacerbated political tensions within schools. Those seeking to enact similar programs should recognize that the very idea of differentiating pay based on performance will likely run up against pervasive school cultures and norms of collaboration and egalitarianism.

- *Pilot testing and evaluation are essential.* From the outset, NYCDOE and UFT leaders intended to implement SPBP on a pilot basis. Resource constraints prevented the planned scale-up within the initial group of schools identified as eligible for this pilot program, providing evaluators with three years to compare outcomes and practices across a set of participating and comparable high-needs control schools. Implementing the program on a small scale and including randomized treatment and control groups for three years (perhaps not intended at the outset) provided valuable information to inform future decisions about an essentially untested policy innovation. Leaders created the opportunity to gather comprehensive data on the implementation and outcomes to assess the merits of continuing and further scaling up this policy. The pilot period and evaluation also provided leaders an opportunity to reexamine the theory of action and assumptions underlying SPBP.

# References


Allen, Mary J., and Wendy M. Yen (1979). *Introduction to Measurement Theory*. Long Grove, IL: Waveland Press, Inc.

Ballou, Dale, and Michael Podgursky (1993). "Teachers' Attitudes Toward Merit Pay: Examining Conventional Wisdom." *Industrial and Labor Relations Review*, 47(1), 50–61.

Ballou, Dale, and Matthew G. Springer (2008). "Achievement Trade-Offs and No Child Left Behind." Working paper. Washington, DC: Urban Institute.

Benjamini, Yoav, and Yosef Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society*, Series B, 57(1), 289–300.

Bloom, Howard S. (1984). "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Review*, 8(2), 225–246.

Booher-Jennings, Jennifer (2005). "Below the Bubble: 'Educational Triage' and the Texas Accountability System." *American Educational Research Journal*, 42(2), 231–268.

Chamberlin, Rosemary, Ted Wragg, Gill Haynes, and Caroline Wragg (2002). "Performance-Related Pay and the Teaching Profession: A Review of the Literature." *Research Papers in Education*, 17(1), 31–49.

Coggshall, Jane G., Amber Ott, Ellen Behrstock, and Molly Lasagna (2009). *Supporting Teacher Effectiveness: The View from Generation Y*. Learning Point Associates and Public Agenda.

Condly, S. J., R. E. Clark, and H. D. Stolovitch (2003). "The Effects of Incentives on Workplace Performance: A Meta-Analytic Review of Research Studies." *Performance Improvement Quarterly*, 16(3), 7–24.

Deci, E. L. (1971). "Intrinsic Motivation, Extrinsic Reinforcement, and Inequity." *Journal of Personality and Social Psychology*, 22(1), 113–120.

Elmore, Richard F. (2004). *School Reform from the Inside Out: Policy, Practice, and Performance*. Cambridge, MA: Harvard Education Press.

Fehr, Ernst, and Simon Gachter (2001). "Do Incentive Contracts Crowd Out Voluntary Cooperation?" Discussion Paper. London: Centre for Economic Policy Research.

Figlio, David N., and Lawrence S. Getzler (2002). "Accountability, Ability and Disability: Gaming the System?" NBER Working Paper Series. Cambridge, MA: National Bureau for Economic Research, October 2002. As of May 13, 2011: http://www.nber.org/papers/w9307

Figlio, David N., and Lawrence W. Kenny (2007). "Individual Teacher Incentives and Student Performance." *Journal of Public Economics*, 91(5–6), 901–914.

Fryer, Roland G. (2010). "Financial Incentives and Student Achievement: Evidence from Randomized Trials." NBER Working Paper Series. Cambridge, MA: National Bureau of Economic Research, April 2010. As of March 21, 2011: http://www.nber.org/papers/w15898

Fryer, Roland G. (2011). "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools." NBER Working Paper Series. Cambridge, MA: National Bureau of Economic Research, March 2011. As of March 21, 2011: http://www.nber.org/papers/w16850

Gini, C. (1921). "Measurement of Inequality of Incomes." *The Economic Journal*, 31(121), 124–126.

Glazerman, Steven, and Allison Seifullah (2010). *An Evaluation of the Teacher Advancement Program in Chicago: Year Two Impact Report*. Washington, DC: Mathematica Policy Research.

Glewwe, P., N. Ilias, and M. Kremer (2004). "Teacher Incentives." handout. Cambridge, MA: Harvard University.

Goodman, Sarena, and Lesley Turner (2011). "Does Whole-School Performance Pay Improve Student Learning? Evidence from the New York City Schools," *Education Next*, 11(2)

Greene, William H. (2007) *Econometric Analysis*. 6th ed. Upper Saddle River, NJ: Prentice Hall.

Hamilton, Laura (2004). "Assessment as a Policy Tool." *Review of Research in Education*, 27, 25–68.

Hamilton, Laura S., Brian M. Stecher, Julie A. Marsh, Jennifer Sloan McCombs, Abby Robyn, Jennifer Russell, Scott Naftel, and Heather Barney (2007). *Standards-Based Accountability Under No Child Left Behind: Experiences of Teachers and Administrators in Three States*. Santa Monica, CA: RAND Corporation, MG-589-EDU. As of March 30, 2011: http://www.rand.org/pubs/monographs/MG589.html

Hannaway, J., and Hamilton, L. (2008). *Performance-Based Accountability Policies: Implications for School and Classroom Practices*. U.S. Department of Education.

Heinrich, Carolyn J. (2007). "False or Fitting Recognition? The Use of High Performance Bonuses in Motivating Organizational Achievements." *Journal of Policy Analysis and Management*, 26(2), 281–304.

Heneman III, Herbert G. (1998). "Assessment of the Motivational Reactions of Teachers to a School-Based Performance Award Program." *Journal of Personnel Evaluation in Education*, 12(1), 43–59.

Heneman III, Herbert G., and Anthony T. Milanowski (1999). "Teachers Attitudes About Teacher Bonuses Under School-Based Performance Award Programs." *Journal of Personnel Evaluation in Education*, 12(4), 327–341.

Jacob, Brian A., and Steven D. Levitt (2003). "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics*, 118, 843–77.

Johnson, S.M. (1986). "Incentives for Teachers: What Motivates, What Matters." *Educational Administration Quarterly*, 22(3), 54–79.

Johnson, Susan Moore, and John P. Papay (2009). *Redesigning Teacher Pay: A System for the Next Generation of Educators*. Washington, DC: Economic Policy Institute.

Kelley, Carolyn (1998). "The Kentucky School-Based Performance Award Program: School-Level Effects." *Educational Policy*, 12(3), 305–324.

Kelley, C. (1999). "The Motivational Impact of School-Based Performance Awards." *Journal of Personnel Evaluation in Education*, 12(4), 309–326.

Kelley, Carolyn J., and Kara Finnigan (2003). "The Effects of Organizational Context on Teacher Expectancy." *Educational Administration Quarterly*, 39(5), 603–634.

Kelley, Carolyn, Herbert Heneman III, and Anthony Milanowski (2002). "Teacher Motivation and School-Based Performance Awards." *Educational Administration Quarterly*, 38(3), 372–401.

Kelley, Carolyn, Odden, A., Milanowski, A., and Heneman III, H. (2000). "The Motivational Effects of School-Based Performance." CPRE Policy Brief, RB-29. Consortium for Policy Research in Education.

Kellor, Ellen M. (2003). "Catching Up with the Vaughn Express: Four Years of Performance Pay and Standards-Based Teacher Evaluation." CPRE-UW Working Paper Series. Madison, WI: Wisconsin Center for Education Research.

Koppich, Julia, and Jessica Rigby (2009). *Alternative Teacher Compensation: A Primer. Policy Analysis for California Education*. Berkeley, CA: University of California Press.

Koretz, D. (2002). "Limitations in the Use of Achievement Tests as Measures of Educators' Productivity." In E. Hanushek, J. Heckman, and D. Neal, eds., *Designing Incentives to Promote Human Capital*. Special issue of *The Journal of Human Resources*, 37(4, fall), 752–777.

Koretz, Daniel, Sheila Barron, Karen Mitchell, and Brian M. Stecher (1996). *The Perceived Effects of the Kentucky Instructional Results Information System* (KIRIS). Santa Monica, CA: RAND Corporation, MR-792-PCT/FF. As of March 30, 2010:
http://www.rand.org/pubs/monograph_reports/MR792.html

Ladd, Helen F., and Douglas L. Lauen (2010) "Status Versus Growth: The Distributional Effects of School Accountability Policies." *Journal of Policy Analysis and Management*. As of ___:
http://www3.interscience.wiley.com/cgi-bin/fulltext/123492093/HTMLSTART. Published Online: Jun 3 2010 10:15AM, DOI: 10.1002/pam.20504

Lipsky, D.B. and Bacharach, S.B. (1983). "The Single Salary Schedule vs. Merit Pay: An Examination of the Debate." *Collective Bargaining Quarterly*, 11(4), 1–11.

Long, J. Scott, and Jeremy Freese (2006). *Regression Models for Categorical Dependent Variables Using Stata*. College Station, TX: StataCorp LP. 2nd Edition.

Lorenz, M. O. (1905). "Methods of Measuring the Concentration of Wealth." *Publications of the American Statistical Association*, 9, 209–219.

Lortie, Dan C. (1975). *Schoolteacher: A Sociological Study*. Chicago, IL: The University of Chicago Press.

Malen, Betty (1999). "On Rewards, Punishments, and Possibilities: Teacher Compensation as an Instrument for Educational Reform." *Journal of Personnel Evaluation in Education*, 12(4), 387–394.

Marsden, David, and Ray Richardson (1994). "Performing for Pay? The Effects of 'Merit Pay' on Motivation in a Public Service." *British Journal of Industrial Relations*, 32(2), 243–261.

McLaughlin, Milbrey Wallin (1987). "Learning from Experience: Lessons from Policy Implementation." *Educational Evaluation and Policy Analysis*, 9(2), 171–178.

Milanoswki, Anthony T. (1999). "Measurement Error or Meaningful Change? The Consistency of School Achievement in Two School-Based Performance Award Programs." *Journal of Personnel Evaluation in Education*, 12(4), 343–363.

Milanowski, A. (2000). School-Based Performance Award Programs and Teacher Motivation. *Journal of Education Finance*, 25(4), 517–544.

Milanowski, A., and Gallagher, A. (2000). "Vaughn Next Century Learning Center Performance Pay Survey School Report." Unpublished manuscript. Madison WI: University of Wisconsin-Madison.

Milkovich, George T., Alexandra K. Wigdor, eds. (1991). *Pay for Performance: Evaluating Performance Appraisal and Merit Pay*. Washington, DC: National Research Council.

Murnane, Richard J., David K. Cohen (1986). "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive." *Harvard Educational Review*, 56(1), 1–17.

National Commission on Excellence in Education (1983). *A Nation At Risk: The Imperative for Educational Reform*. Washington, DC: U.S. Department of Education.

National Institute for Excellence in Teaching, TAP: The System for Teacher and Student Advantage, website, 2011. As of December 31, 2010:
TAP webpage, http://www.tapsystem.org

Neal, Derek, and Diane Whitmore Schanzenbach (2010). "Left Behind by Design: Proficiency Counts and Test-Based Accountability." *Review of Economics and Statistics*, 92(2), 263–283.

New York City Department of Education (undated). "Quick Reference Guide."

——— (2007) Mayor, Chancellor, UFT Announce Plan to Reward Teachers at Schools that Raise Student Achievement. Press release.

——— (2009a). 2009 School-Wide Performance Bonus Program Training. Powerpoint presentation.

——— (2009b). Educator Guide: The New York City Progress Report, Elementary/Middle Schools.

——— (2009c). Elementary, Middle, High, and Schools for Transfer Students Receiving 2007/08 Progress Reports, Excel file, January 13, 2009. As of December 18, 2010:
http://schools.nyc.gov/Accountability/tools/report/Prior+Year+Progress+Report+Results.htm

——— (2009d). Elementary, Middle, High, and Schools for Transfer Students Schools Receiving 2006/07 Progress Report Grades, Excel file, January 14, 2009. As of December 18, 2010:
http://schools.nyc.gov/Accountability/tools/report/Prior+Year+Progress+Report+Results.htm

——— (2009e). Chancellor Klein Announces Record-High Participation in Annual Public School Survey. News release June 24, 2009.

——— (2010a). 2009–2010 High School Progress Report Results, ca. 2010. As of December 18, 2010:
http://schools.nyc.gov/Accountability/tools/report/default.htm

——— (2010b). Elementary, Middle, High, and Schools for Transfer Students Receiving 2008/09 Progress Reports, Excel file, March 23, 2010. As of December 18, 2010:
http://schools.nyc.gov/Accountability/tools/report/Prior+Year+Progress+Report+Results.htm

——— (2010c). "Chancellor Klein Releases 2010 Progress Reports for Schools Serving Students in Grades Kindergarten–8," news release, September 9, 2010. As of May 9, 2011:
http://schools.nyc.gov/Offices/mediarelations/NewsandSpeeches/2010-2011/2010ProgressReportsAnnouncement.htm

——— (2010d). Elementary, Middle, Early Childhood, and District 75 Schools Receiving 2009/10 Progress Reports, October 5, 2010. As of December 18, 2010:
http://schools.nyc.gov/Accountability/tools/report/Prior+Year+Progress+Report+Results.htm

——— (2011a), "Transfer Schools," website. As of January 13, 2011:
http://schools.nyc.gov/ChoicesEnrollment/AlternativesHS/TransferHS/default.htm

——— (2011b), "Progress Report," website. As of March 30, 2011:
http://schools.nyc.gov/Accountability/tools/report/default.htm

——— (2011c), "Progress Report: Prior Year Progress Report Results," website. As of March 30, 2011:
http://schools.nyc.gov/Accountability/tools/report/Prior+Year+Progress+Report+Results.htm

New York City Department of Education and United Federation of Teachers (2010). "School-Wide Performance Bonus Program (SPBP) 2009–2010: Frequently Asked Questions (FAQs)"

NYCDOE—See New York City Department of Education.

Odden, A. (2001). "Rewarding Expertise." *Education Matters*, 1(1), 16–24.

Odden, Allan, and Carolyn Kelley (1997). *Paying Teachers for What They Know and Do: New and Smarter Compensation Strategies to Improve Schools*. Thousand Oaks, CA: Corwin Press.

Padilla, C., Skolnik, H., Lopez-Torkos, A., Woodworth, K., Lash, A., Shields, P. M., David, J.L. (2006). Title I Accountability and School Improvement from 2001 to 2004. Washington, DC: U.S. Department of Education.

Partnership for New York City (2011), "About," web page. As of March 28, 2011:
http://nycp.org/about.html

Pearce, Jone L., and James L Perry (1983). "Federal Merit Pay: A Longitudinal Analysis." *Public Administration Review*, 43(4), 315–325.

Pechthalt, J. (2007). "No Merit to Merit Pay." *United Teacher*, 37(3). As of March 9, 2011:
http://www.utla.net/node/930

Podgursky, Michael J., and Matthew G Springer (2007). "Teacher Performance Pay: A Review." *Journal of Policy Analysis and Management*, 26(4), 909–949.

Reback, Randall (2008). "Teaching to the Rating: School Accountability and the Distribution of Student Achievement." *Journal of Public Economics*, 92(5–6), 1394–1415.

Riccio, James, Nadine Dechausay, David Greenberg, Cynthia Miller, Zawadi Rucks, and Nandita Verma (2010). *Toward Reduced Poverty Across Generations: Early Findings from New York City's Conditional Cash Transfer Program*. New York, NY: MDRC.

Shadish, William R., Thomas D Cook, and Donald Thomas Campbell (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.

Snedecor, G.W. and Ames, W.G.C. (1989). *Statistical Methods*, 8th ed. Iowa State University Press.

Solmon, L. (2006). "The Case for Merit Pay." In P. Peterson (ed.), *Choice and Competition in American Education*. Lanham, MD: Rowman and Littlefield Publishers.

Springer, Matthew G. (2008). "The Influence of an NCLB Accountability Plan on the Distribution of Student Test Score Gains." *Economics of Education Review*, 27(5), 556–563.

———, ed. (2009) *Performance Incentives: Their Growing Impact on American K–12 Education*. Washington, DC: Brookings Institution Press.

Springer, Matthew G., and Balch, R. (2009). "Design Components of Incentive Pay Programs in the Education Sector." unpublished manuscript.

Springer, M. G., Ballou, D., Burns, S.F., Hamilton, L., Le, V., Lockwood, J.R., McCaffrey, D., Pane, J., and Stecher, B. (2010c). *Teacher Pay for Performance: Experimental Evidence from Round Rock's Project on Incentives in Teaching*. Nashville, TN: National Center on Performance Incentives.

Springer, Matthew G., Dale Ballou, and Art Peng (2008). "The Impact of Teacher Advancement Program on Student Test Score Gains: Findings from an Independent Appraisal." Working paper. Nashville, TN: National Center on Performance Incentives.

Springer, Matthew G., Laura Hamilton, Daniel F. McCaffrey, Dale Ballou, Vi-Nhuan Le, Matthew Pepper, J. R. Lockwood, and Brian M. Stecher. (2010a). *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT)*. Nashville, TN: National Center on Performance Incentives at Vanderbilt University.

Springer, Matthew G., Jessica L. Lewis, Michael J. Podgursky, Mark W. Ehlert, Timothy J Gronberg;, Laura S. Hamilton, Dennis W. Jansen, Brian M. Stecher, Lori L. Taylor, Omar S. Lopez, and Art Peng (2009a). *Texas Educator Excellence Grant (TEEG) Program: Year Three Evaluation Report*. Nashville, TN: National Center on Performance Incentives.

Springer, Matthew G., Jessica L. Lewis, Michael J. Podgursky, Mark W. Ehlert, Lori L. Taylor, Omar S. Lopez, and Art Peng (2009b). *Governor's Educator Excellence Grant (GEEG) Program: Year Three Evaluation Report*. Nashville, TN: National Center on Performance Incentives.

Springer, Matthew G., and Marcus A. Winters (2009). "New York City's School-Wide Bonus Pay Program: Early Evidence from a Randomized Trial." Working paper. National Center on Performance Incentives, Vanderbilt University.

Stecher, Brian M., Scott Epstein, Laura S. Hamilton, Julie A. Marsh, Abby Robyn, Jennifer Sloan McCombs, Jennifer Russell, and Scott Naftel (2008), *Pain and Gain: Implementing No Child Left Behind in Three States, 2004–2006*. Santa Monica, CA: RAND Corporation, MG-784-NSF. As of March 30, 2011: http://www.rand.org/pubs/monographs/MG784.html

Stronge, James H., Christopher R. Gareis, and Catherine A. Little (2006). *Teacher Pay and Teacher Quality: Attracting, Developing and Retaining the Best Teacher*s. Thousand Oaks: Corwin Books.

Taylor, Lori L., and Matthew G. Springer (2009). "Optimal Incentives for Public Sector Workers: The Case of Teacher-Designed Incentive Pay in Texas." Working paper. Nashville, TN: National Center on Performance Incentives.

Taylor, L., Springer, M.G., and Ehlert, M. (2009). "Teacher-Designed Performance-Pay Plans in Texas." in *Performance Incentives: Their Growing Impact on American K–12 Education*. Washington, DC: Brookings Institution Press. pp. 191–224.

United Federation of Teachers (2011), "Union Basics," web page. As of February 12, 2011: http://www.uft.org/who-we-are/union-basics

U.S. Department of Education (2009). *Race to the Top Program Executive Summary*. Washington, DC. As of May 25, 2011: http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

Vroom, V. H. (1964). *Work and Motivation*. New York: Wiley.