Learning from Pay-for-Performance Programs:

*Insights from New York City’s Performance Bonus Experiment*

Julie Marsh
Associate Professor
University of Southern California
Outline

- NYC Schoolwide Performance Bonus Program (SPBP)
  - Relevance to your work
  - Overview of program
- The Study
  - Study overview
  - Research process
- Key Findings
- Questions/Discussion
How is this case relevant to your work?

- Pay-for-performance programs in education are similar to results-based financing (RBF) for health
  - Part of broader shifts in public sector accountability
  - Similar goals & mechanisms

- Challenges to conducting research in NYC mirror those in developing countries
  - Limited capacity, data, access
  - Complexity of embedded contexts

- Our study methods match those called for in your research of RBF
Design of NYC SPBP

- Jointly implemented by NYCDOE & United Federation of Teacher (UFT)
- Three-year voluntary program in high-needs K-12 schools
- Schools meeting annual targets based on Progress Reports earn school-level bonuses worth:
  - $3,000 per full-time UFT-represented staff (meet 100% target)
  - $1,500 per full-time UFT-represented staff (meet at least 75% target)
  - Principals earn $7,000 (full bonus)/$3,500 (partial): APs get half
- Progress Report scores determined by:
  - Student growth on state tests & performance relative to other schools
  - High school graduation & progress to graduation (credits, Regents Exams completion & pass rates)
  - Conditions for learning (attendance; survey data)
- 4-person school Compensation Committee (CC) decides distribution
Rollout of NYC SPBP

- In Year 1 (2007-2008), 427 schools identified & about ½ randomly selected for opportunity to participate …

<table>
<thead>
<tr>
<th></th>
<th># Schools Invited</th>
<th>Voted No</th>
<th>Voted Yes</th>
<th>Withdrew or closed</th>
<th>Added</th>
<th># Schools Participating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>234</td>
<td>32</td>
<td>199</td>
<td>-3</td>
<td>+6</td>
<td>205</td>
</tr>
<tr>
<td>Year 2</td>
<td>205</td>
<td>4</td>
<td>197</td>
<td>-4</td>
<td>+1</td>
<td>198</td>
</tr>
<tr>
<td>Year 3</td>
<td>198</td>
<td>2</td>
<td>196</td>
<td></td>
<td></td>
<td>196</td>
</tr>
</tbody>
</table>

- Program distributed $55.5 million over 3 years:
  - Year 1: 62% of schools won bonuses – $20.8 million
  - Year 2: 84% of schools won bonuses - $30.6 million
  - Year 3: 13% of schools won bonuses - $4.1 million
    (state raised proficiency threshold)
Outline

- NYC Schoolwide Performance Bonus Program (SPBP)
  - Relevance to your work
  - Overview of program
- The Study
  - Study overview
  - Research process
- Key Findings
- Questions/Discussion
Study Overview

△ Research Questions

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

△ Design

- Takes advantage of program random assignment
- Draws on qualitative & quantitative data
Research Process

1. Conceptual Framework
2. Planning & Data Collection
3. Data Analysis
Step 1: Conceptual Framework

- Understanding Program’s Design & Theory of Action

- DOE-UFT Implementation
  - Lottery
  - Progress Report & targets
  - Information & communication
  - Technical Assistance
  - Tools
  - Funding

- School Implementation
  - Vote
  - Committee process

- SPBP Theory of Action
  - Improved Motivation (anticipatory) & Morale (post-receipt)
  - Improved School & Classroom Practices
  - Improved Student Achievement
  - Improved Staff Recruitment & Retention
  - Improved Staff Collaboration & Cooperation

- BONUS
  - financial incentive
  - group incentive

- Developing & Refining the Conceptual Framework …
Step 1: Conceptual Framework

DOE-UFT Implementation
- Lottery
- Progress Report & targets
- Information & communication
- Technical Assistance
- Tools
- Funding

School Implementation
- Vote
- Committee process

SPBP Theory of Action
- BONUS
  - financial incentive
  - group incentive
- Improved Motivation (anticipatory) & Morale (post-receipt)
- Improved Staff Collaboration & Cooperation
- Improved Staff Recruitment & Retention
- Improved School & Classroom Practices
- Improved Student Achievement

Community, School, Classroom Context & Mediating Factors
School size, leadership, capacity, professional culture, other accountability policies
Individual opinions, attitudes, understanding, buy-in, expectancy, other characteristics
## Step 2: Planning & Data Collection

<table>
<thead>
<tr>
<th>Key Questions &amp; Constructs</th>
<th>Data Sources</th>
</tr>
</thead>
</table>
| **System- & school-level implementation** | • Are they doing what they were supposed to do? (“fidelity” measures)  
• What is the nature & quality of these activities? | • Leader interviews  
• Docs/Admin data  
• Surveys & Case studies |
| **Context & mediating variables** | • To what extent are conditions predicted by theory/research in place?  
- *Understanding, buy-in, valence, expectancy, perceived fairness*  
• What other factors are hindering or facilitating implementation? | • Surveys (Treat)  
• Case studies  
• Leader interviews |
| **Intermediate outcomes** | • Is there evidence of expected changes on intermediate outcomes?  
- *Motivation, morale, collaboration, school & classroom practices* | • Surveys (Treat & Control)  
• Case studies |
| **Student outcomes** | • Is there evidence of program effects on student outcomes?  
- *Student test results, Progress Reports*  
• Are there relationships between implementation & outcomes? | • Achievement data  
• Demographic data  
• Admin data |
# Data Sources: Leader Interviews & Docs (qualitative)

<table>
<thead>
<tr>
<th></th>
<th>Leader Interviews</th>
<th>Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WHO</strong></td>
<td>Key designers, implementers, funders (n=17)</td>
<td>Sponsor organizations, media</td>
</tr>
<tr>
<td><strong>HOW</strong></td>
<td>Semi-structured interviews: telephone &amp; in person</td>
<td>Web searches, requests</td>
</tr>
<tr>
<td><strong>WHEN</strong></td>
<td>Annually (Spring)</td>
<td>Ongoing</td>
</tr>
<tr>
<td><strong>WHAT</strong></td>
<td>• Program history, goals, mechanisms</td>
<td>• Program design</td>
</tr>
<tr>
<td></td>
<td>• Perceptions about implementation, mediating factors</td>
<td>• How program was communicated to participants, public</td>
</tr>
<tr>
<td></td>
<td>• Changes made to program design over time</td>
<td>• Broader context (e.g., other policies, conditions)</td>
</tr>
<tr>
<td></td>
<td>• Lessons learned</td>
<td></td>
</tr>
</tbody>
</table>
## Data Sources: Surveys (quantitative)

<table>
<thead>
<tr>
<th></th>
<th>CC Member Surveys</th>
<th>Teacher Surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WHO</strong></td>
<td>All members in all participating schools</td>
<td>Stratified random sample of tested &amp; non-tested grades/subjects in treatment &amp; control schools</td>
</tr>
<tr>
<td>2009: n=537, 68% response</td>
<td></td>
<td>n=1,532: 57% response</td>
</tr>
<tr>
<td>2010: n=561, 72% response</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HOW</strong></td>
<td>Paper &amp; web-based versions</td>
<td>Web-based, gift cards</td>
</tr>
<tr>
<td><strong>WHEN</strong></td>
<td>Annually (end of school year)</td>
<td>End of school year 2010</td>
</tr>
<tr>
<td><strong>WHAT</strong></td>
<td>• Perceptions about design, implementation, effects</td>
<td>• Both groups: intermediate outcomes (instructional practices, motivation)</td>
</tr>
<tr>
<td></td>
<td>• Mediating factors</td>
<td>• Treatment only: perceptions about implementation, context/mediating factors</td>
</tr>
</tbody>
</table>
# Data Sources: Case Studies (qualitative)

<table>
<thead>
<tr>
<th>Case Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WHO</strong></td>
</tr>
<tr>
<td>• Purposive sample: 7 schools (2009); 7 different schools (2010)</td>
</tr>
<tr>
<td>• 7 elementary, 3 middle, 4 high: some won &amp; loss bonus</td>
</tr>
<tr>
<td>• Oversampling of schools with differentiated distribution plans</td>
</tr>
<tr>
<td>• Interviews with 131 administrators, teachers, non-teacher staff</td>
</tr>
<tr>
<td><strong>HOW</strong></td>
</tr>
<tr>
<td>In-person visits; semi-structured interviews</td>
</tr>
<tr>
<td><strong>WHEN</strong></td>
</tr>
<tr>
<td>Annually (end of school year)</td>
</tr>
<tr>
<td><strong>WHAT</strong></td>
</tr>
<tr>
<td>• Perceptions about design, implementation, effects</td>
</tr>
<tr>
<td>• Mediating factors</td>
</tr>
<tr>
<td>• Unmet needs, suggestions for program improvement</td>
</tr>
</tbody>
</table>
Data Sources: Administrative Data (quantitative)

- **Administrative data (lottery, voting, distribution plans)**
- **School-level data**
  - Demographic, attendance, disciplinary data
  - AYP status/ratings
  - Progress Report scores & grades
- **Student-level data**
  - Student biographical data
  - Student achievement test scores
    - All three program years for elementary/middle schools
    - First two program years for high schools
**SURVEY:** How well do you understand the following aspects of the Schoolwide Performance Bonus Program this year?

<table>
<thead>
<tr>
<th>Question</th>
<th>Not very well</th>
<th>Somewhat well</th>
<th>Very well</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Criteria for a school to receive a full bonus</td>
<td>☐ 1</td>
<td>☐ 2</td>
<td>☐ 3</td>
</tr>
<tr>
<td>b. Which staff are eligible to receive a share of the schoolwide bonus</td>
<td>☐ 1</td>
<td>☐ 2</td>
<td>☐ 3</td>
</tr>
<tr>
<td>c. The target our school needs to reach in order to earn a bonus this year</td>
<td>☐ 1</td>
<td>☐ 2</td>
<td>☐ 3</td>
</tr>
<tr>
<td>d. The amount of money our school would receive if we met 100% of our target</td>
<td>☐ 1</td>
<td>☐ 2</td>
<td>☐ 3</td>
</tr>
</tbody>
</table>

**INTERVIEW:**
- Do you have a good understanding of how the program works?
- Do you know how the bonuses were awarded to schools last year?
- What are the criteria?
- How and from whom have you received information about the program? Has the information been clear and easy to understand?
Instrument Design: Measures of Valence

For those who received the bonus:

- **SURVEY**: To what extent do you agree or disagree … [4-point scale]
  - I was satisfied with the amount I received
  - After taxes, the bonus amount I received felt insignificant

- **INTERVIEW**:
  - Were you satisfied with the amount?
  - When we look at the formal program documents it states that the average per person award amount should be “sufficiently substantial to make a material difference” — do you see it that way?
  - Why or why not?
Instrument Design: Measures of Motivation

SURVEY: To what extent do you agree or disagree … [4-point scale]
- Receiving my bonus share was a nice acknowledgement of my effort but did not influence the way I perform my job this year
- After receiving the bonus I worked harder this year OR
- Not receiving a bonus reduced my motivation to work harder this year
- Not receiving a bonus energized me to improve my teaching this year

INTERVIEW:
- How has participation in the SPBP affected your motivation? Your level of job stress?
- To what extent has receiving the bonus affected your desire to remain or not remain in the school? Has it affected you in any other way?
Step 3: Analyses - Implementation

- To understand how the program was implemented:
  - Coded interview notes & transcripts & documents along dimensions of conceptual framework
  - Developed individual school analytic memoranda
Step 3: Analyses – Implementation & Mediating Factors

To understand how the program was implemented:

- Coded interview notes & transcripts & documents along dimensions of conceptual framework
- Developed individual school analytic memoranda
- Examined frequencies &

Analytic Memo for School A

- Context (size, population, community, culture)
- Implementation
  - Level of support for participation
  - Experience of CC members
  - Distribution plan
  - Receipt of bonus
- Communication & Technical Assistance
  - Level of understanding
    - 4 of the 6 teachers did not know the amount of funding the school would receive
    - “I don’t really know. Maybe $500?” (5th grade T)
- Data & Tools
- Communication to School Stakeholders
- Perceived Effects (on recruitment, staff interactions, individual affective & behavioral)
- General Attitudes & Opinions about SPBC
- Advice for Improving the Program
Step 3: Analyses – Implementation & Mediating Factors

- To understand how the program was implemented:
  - Coded interview notes & transcripts & documents along dimensions of conceptual framework
  - Developed individual school analytic memoranda
  - Examined frequencies & cross-tabulations of survey data to explore relationships
  - Developed cross-case analytic memoranda combining case study & survey data
Step 3: Analyses – Implementation & Mediating Factors

To understand how the program was implemented:
- Coded interview notes & transcripts & documents along dimensions of conceptual framework
- Developed individual school analytic memoranda
- Examined frequencies & cross-tabulations of survey data to explore relationships
- Developed cross-case analytic memoranda combining case study & survey data

Cross-Case Analytic Memo: Communication & Understanding

• **Leader interviews**
  - Leaders from DOE & UFT acknowledge communication problems in first year
  - Leaders & schools report improvements in later years

• **Surveys**
  - Majority of CC members report receiving sufficient guidance
  - CC members report improvements in understanding of program elements in year 2: slightly less in year 3
  - 84% of teachers report strong understanding of school’s progress report & contributing factors
  - Yet sizeable minority report not understanding several program elements
  - Reported understanding higher in schools earning bonus

• **Cases**
  - Widespread misunderstandings conveyed in interviews about x, y ..
  - In almost every school misperceptions were conveyed
Step 3: Analyses - Intermediate Outcomes

- To understand program effects on teachers:
  - Intent to Treat (ITT) Analysis: Compared self-reported attitudes & behaviors of teachers in schools randomly assigned to SPBP with those in schools assigned to control
  - Created composite measures/scales from survey responses
  - Used mixed-effects two-level hierarchical model to examine whether treatment teachers differed significantly from control teachers (scales)
  - Applied ordinal logit & logit regression model (individual survey items)
  - Adjusted for multiple comparisons (Benjamini & Hochberg, 1985)
  - Illustrated with examples from case studies
To understand program effects on student test scores:

- ITT Analysis

- Ordinary least squares regression analysis with cluster adjusted standard errors at school level

- Dependent variable:
  - Standardized z-scores & scale scores in ELA & math (elementary, middle, K-8)
  - Separate effects estimated for 6 Regents Exams (2 required & most commonly taken) (high schools)

- Controls: student & school characteristics, incl. prior achievement
NYC Schoolwide Performance Bonus Program (SPBP)
- Relevance to your work
- Overview of program

The Study
- Study overview
- Research process

Key Findings

Questions/Discussion
## Overview of Findings

| Student outcomes | SPBP did not improve student achievement at any grade level: (elementary, middle, k-8) or affect Progress Report performance  
|                  | No stat. sign. relationship between achievement & bonus differentiation |
| Intermediate outcomes | SPBP did not affect teacher reported attitudes, perceptions, & behaviors  
|                      | Teachers in control schools were just as likely as those from treatment schools to report undertaking efforts to help their school achieve a high Progress Report grade |
| System- & school-level implementation | Basic procedures were enacted as planned  
|                                    | Communications followed the SPBP design: early problems corrected over time  
|                                    | CC process implemented smoothly: Some had difficulty with decision-making  
|                                    | Majority of CCs developed nearly egalitarian plans |
| Context & mediating variables | Key supporting conditions were lacking (Understanding, buy-in, valence)  
|                               | Other contextual factors may have limited bonus’ motivational power  
|                               | All schools faced significant pressure to perform well on the same measures incentivized by SPBP  
|                               | Other accountability pressures & intrinsic motivation were often perceived to be more salient than the bonus |
Outline

- NYC Schoolwide Performance Bonus Program (SPBP)
  - Relevance to your work
  - Overview of program
- The Study
  - Study overview
  - Research process
- Key Findings
- Questions/Discussion
Additional slides on findings
SPBP Did Not Improve Student Achievement at Any Grade Level: Elementary, Middle, K-8

<table>
<thead>
<tr>
<th>Subject</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Language Arts (ELA)</td>
<td>--</td>
<td>-0.03</td>
<td>-0.03</td>
</tr>
<tr>
<td>(std. error)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Mathematics</td>
<td>-0.02</td>
<td>-0.04</td>
<td>-0.06*</td>
</tr>
<tr>
<td>(std. error)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.03)</td>
</tr>
</tbody>
</table>

* Significant at the 0.05 level
SPBP Did Not Affect Teacher Reported Attitudes, Perceptions, & Behaviors

- Virtually no statistically significant differences between reported practices & opinions of teachers in treatment & control group schools

  - Instructional practices (e.g., using test data to guide instruction, aligning instruction with standards)
  - Effort (e.g., time spent on work outside regular school hours)
  - Participation in professional development
  - Collaboration with other teachers or administrators
  - Mobility (e.g., plans to return to their school next year)
  - Attitudes (e.g., support for test-based rewards)
SPBP Did Not Affect Teacher Reported Attitudes, Perceptions, & Behaviors: Effort

- Time spent on work outside regular school hours (hours per week):
  - Treatment: 12.8
  - Control: 13.17

- Overall hours on professional development this school year:
  - Treatment: 48.48
  - Control: 49.95

Legend:
- treatment
- control
SPBP Did Not Affect Teacher Reported Attitudes, Perceptions, & Behaviors: 

<table>
<thead>
<tr>
<th>Collaboration</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaboration with other teachers</td>
<td>3.43</td>
<td>3.47</td>
</tr>
<tr>
<td>Collaboration with administrators</td>
<td>3.02</td>
<td>3.05</td>
</tr>
<tr>
<td>Collaboration with whole staff</td>
<td>3.53</td>
<td>3.5</td>
</tr>
</tbody>
</table>
What Explains the Lack of Positive Effects?

- Implementation findings provide some insights, but do not provide a complete explanation.
- Let’s start with the implementation results …
Basic Procedures & Processes Were Implemented As Planned

- NYCDOE & UFT generally implemented SPBP as intended
  - Votes occurred
  - CCs were formed, generally followed guidelines & operated without outside interference
  - Distribution plans were submitted
  - Overall distribution process reported to run smoothly: Few reported problems & none reached level of official appeal
Yet Implementation Did Not Establish All of The Supporting Conditions

- **Understanding**
  - Teachers reported being aware of the program & supportive of it
  - However, more than 1/3 did not understand targets school needed to reach & bonus amounts

- **Buy-In**
  - Majority of teachers (84%) & CC members (74%) felt bonus criteria relied too heavily on student test scores

- **Valence**
  - Majority of teachers expressed strong desire to win bonus
  - 73% of teachers in bonus schools reported that after taxes the amount seemed insignificant
  - Almost ½ indicated bonus was not large enough to motivate extra effort
  - Many case study respondents viewed bonus as reward for their usual efforts, not as incentive for changing behavior
Other Contextual Factors May Have Limited The Motivational Power of The Bonus

- All schools faced significant pressure to perform well on the same measures incentivized by SPBP.

- Teachers in control schools were just as likely as those from treatment schools to report undertaking efforts to help their school achieve a high Progress Report grade:
  - Devoting effort to improve student attendance
  - Seeking PD opportunities to improve their practice
  - Working with students to set & monitor goals

- Other accountability pressures & intrinsic motivation were often perceived to be more salient than the bonus …
<table>
<thead>
<tr>
<th>Motivating Factor</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seeing my students learn new skills and knowledge</td>
<td>99</td>
</tr>
<tr>
<td>Seeing my students improve their performance</td>
<td>99</td>
</tr>
<tr>
<td>Seeing myself improve as a teacher</td>
<td>99</td>
</tr>
<tr>
<td>Achieving or exceeding our school’s Adequate Yearly Progress target</td>
<td>77</td>
</tr>
<tr>
<td>Receiving a high Quality Review score</td>
<td>76</td>
</tr>
<tr>
<td>Achieving a high Progress Report score/grade</td>
<td>75</td>
</tr>
<tr>
<td>Our school being designated for district or state intervention</td>
<td>64</td>
</tr>
<tr>
<td>Receiving a financial bonus for meeting school goals</td>
<td>64</td>
</tr>
<tr>
<td>Gaining recognition from my school colleagues for helping to meet our school goals</td>
<td>64</td>
</tr>
<tr>
<td>Gaining recognition from parents, the media, and/or district leaders for meeting our school goals</td>
<td>62</td>
</tr>
<tr>
<td>Our school being closed or phased out</td>
<td>58</td>
</tr>
<tr>
<td>Receiving public criticism due to not achieving our school goals</td>
<td>44</td>
</tr>
</tbody>
</table>
There Are Other Possible Explanations That Our Data Do Not Allow Us To Substantiate

- Motivational potential of SPBP was not realized due to the widespread adoption of egalitarian plans
  - If more schools had differentiated payout based on individual performance there may have been greater effects

- SPBP was incomplete & did not provide needed capacity-building & resources
  - Bonuses may have motivated educators to improve but they lacked the knowledge, leadership, social capital, materials, etc. to bring about improvement

- The real effects of SPBP may be realized long term through labor market effects
  - If SPBP had been implemented for a longer period of time it may have helped to attract & retain high-quality educators to SPBP schools & incentivized less-qualified individuals to exit