MANAGER AND EMPLOYEE BEHAVIORS IN RESPONSE TO PERFORMANCE SIGNALS IN PUBLIC ORGANIZATIONS

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

TAEK KYU KIM

Dr. Weijie Wang, Dissertation Co-supervisor

Dr. Lael Keiser, Dissertation Co-supervisor

July 2023

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

Manager and Employee Behaviors in Response to Performance Signals in Public Organizations

presented by Taek Kyu Kim,

a candidate for the degree of doctor of philosophy in Public Affairs,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Lael Keiser

Professor Weijie Wang

Professor David Switzer

Professor Bradley Curs

ACKNOWLEDGEMENTS

One question that drew me into academia is “Why does the public sector seem more inefficient than other sectors?” I was interested in public performance management, which covers evaluation, analyses, and utilization of outcomes from employee and organization activities. Through studying in undergraduate, master, and Ph.D. courses, I now deeply understand the uniqueness of the public sector. However, I still believe there are unexamined ways to improve public service quality and citizen satisfaction through performance management. My study on performance information feedback started with this interest.

Despite my passion and a keen interest in my study, I should fully acknowledge that this study could not be possible without people who supported me in my path. Throughout my Ph.D. student life and dissertation writing, it was a great fortune to work together with Weijie Wang who is my mentor and my dissertation co-chair. I have thought of him as my eldest brother. He has provided unlimited support and advice throughout my life as a Ph.D. student, and I could expand my capacity as a researcher by conducting several research projects with him. I am grateful for Lael Keiser who is my dissertation co-chair. Her invaluable intellectual guidance helped me to develop my theory and enhance the completeness of my dissertation. Other committee members, David Switzer and Bradley Curs, also hugely contribute to this dissertation through patient reading and constructive comments. I am also thankful to Irma Arteaga for helping me to keep financial independence during my Ph.D. Her recommendation for the G. Ellsworth Huggins

Fellowship allowed me to have abundant financial resources for 5 years, which made a favorable environment to only focus on study.

I deeply appreciate Sounman Hong (Yonsei University), Hah-zoong Song, Dohan Kim (Kyung Hee University), and Ju-Kyong Kim (Jeju National University) for their support and encouragement. Sounman Hong encourage me to have a bigger dream as a public management scholar and provided invaluable support to study in the United States. I could not have discovered my interest in public administration and policy without the experience as a research assistant given by Ju-Kyong Kim; and couldn't have decided to enter Yonsei University for a master's degree without advice from Hah-zoong Song and Dohan Kim. I want to thank Brian Kisida for giving me an opportunity to explore other topics in public management and enlarge my skillsets. I am also thankful to my fellow colleagues, especially Hoyeon Lee and Youngrok Kim, for fueling my passion for research and stimulating research ideas.

Finally, I can reach this stage thanks to the unlimited support and love from my parents and younger brother. They were great counselors and supporters through the ups and downs throughout my whole life. My lovely girlfriend has always been cheerful and supportive on my path. I am deeply indebted to her dedication, and it is undeniable that it made a difference in my dissertation and job-searching process. I would like to dedicate my dissertation to all of them.

TABLE OF CONTENTS

Manager and Employee Behaviors in Response to Performance Signals in Public Organizations

Taek Kyu Kim

Dr. Weijie Wang, Dissertation Co-supervisor

Dr. Lael Keiser, Dissertation Co-supervisor

ABSTRACT

This dissertation explores the diverse behaviors of managers and employees in response to performance signals in public organizations. Performance outcomes provide crucial information for actors related to public organizations such as politicians, public officials, and citizens. Previous literature has discussed performance information as a management tool for holding public officials accountable, facilitating organizational learning, and monitoring the activities of public organizations. Nonetheless, the research has been limited to explaining how performance outcomes affect manager and employee decision-making, including organization strategy, goal setting, and even unethical behaviors. This dissertation specifically focuses on manager decision-making on co-production and effort substitution and employee decision-making on employee fraud. Each of these concepts

regards organizational efficiency and public values differently. This dissertation, hence, is laid out into three separate but inter-linked chapters, individually focusing on each of the important concepts in public management.

Chapter 1 is motivated by the following question: Do organizations engage in co-production in response to performance signals? Co-production is a set of activities in which public organization actors and stakeholders work together for public service delivery; it positively impacts reducing costs and improving service quality. However, co-production may be risky because it causes unintended effects such that it complicates the delivery process by empowering external actors. Given the benefits and risks, would organizations deliberately increase effort for co-production under certain circumstances, for example, when they are underperforming? To examine the research question, this study employs a local polynomial regression discontinuity (RD) design based on school performance data and ratings from School Quality Reports in New York City public schools. The findings show that public organizations are more likely to increase the degree of co-production in response to performance shortfall.

If Chapter 1 suggests co-production as a potential strategy based on the assumption of a single-goal environment, in Chapter 2, this dissertation investigates the management strategy in response to performance shortfalls in a multiple-goal environment. Specifically, Chapter 2 focuses on effort substitution, which is an unintended effect of goal prioritization. While goal prioritization leads to performance improvement, it could also cause public managers to deliberately pay less attention to low-stake goals in their performance metrics. It narrowly focuses on efficiency and may harm equity and democracy. This chapter therefore examines whether there is effort substitution towards high-stake goals in public

organizations triggered by negative performance feedback. Similar to the previous chapter, evidence of this chapter comes from a local polynomial regression discontinuity (RD) design based on New York City public school data. Empirical evidence from this study presents that low performing organizations prioritize overall performance outcomes but pay less attention to low-stakes goals.

Different from other chapters focusing on managerial behaviors, Chapter 3 offers a novel perspective on the relationship between performance outcomes and employee behaviors. This chapter investigates employee fraud, which is different from organizational cheating in that it is the employee behavior to gain financial advantages for personal enrichment by misusing organizational resources. Although previous public management research discussed organizational cheating related to performance, what is still relatively unknown is how performance outcomes affect employee fraud. To test this relationship, this study links the data from performance evaluation and integrity assessment for South Korean government corporations. Global polynomial regression discontinuity design (RDD) was employed to estimate the impact of performance shortfalls on employee fraud. The results show that employees in low-performing public corporates are more likely to abuse operating expenses, benefit allowance, and project expenses than employees in relatively high-performing organizations.

In conclusion, the level of past performance is highly influential for manager and employee behaviors. This possibly provides an important reference point for subsequent behaviors as it contributes to the burgeoning performance management research by showing how performance signals affect subsequent managers' and employees' decision-making in diverse contexts. It also offers insight into novel behaviors of managers and

employees in response to performance signals; it connects the performance management research and literature on co-production, goal prioritization, and counterproductive work behaviors. Finally, this dissertation provides an opportunity to discuss problemistic search and accountability mechanism as a theoretical framework to explain the reasons for behaviors.

Chapter 1: How Does Performance Feedback Affect Co-production?

Introduction

Modern public organizations are expected to achieve high service performance with limited resources (Gazley et al., 2020; Nabatchi et al., 2017). Although several crises such as the 2008 financial crisis and the COVID-19 pandemic caused fiscal and other resource constraints for many public services, public organizations still are expected to provide quality services to citizens (Kiewiet & McCubbins, 2014). Furthermore, new management regimes including New Public Governance (NPG) and Open Government (OG) require public organizations to invite the citizens into the process of producing public service (Schmidthuber et al., 2021). These current environments increase the necessity of co-production in the public sector.

Co-production, which indicates activities in which public organization actors and stakeholders work together for public service delivery (Clark et al., 2013; M. Jakobsen & Andersen, 2013; Nabatchi et al., 2017; Ostrom, 1972; Pestoff, 2006; Voorberg et al., 2015), is not a novel concept in public management. Elinor Ostrom and her colleagues introduced the concept in public management studies in the 1970s, and the concept become popularized in 1980s (e.g. Ostrom 1972; Ostrom and Ostrom 1977; Percy 1978; Parks et al. 1981). Co-production is highlighted in public management studies again in the recent years as a way to adapt to current public management environments (Gazley et al., 2020; Nabatchi et al., 2017). Besides the literature examining the effect of co-production, previous literature on co-production has investigated the specific conditions under which

organizations engage in co-production. These conditions cover public organization factors, such as organizational structure and the attitudes of public officials (Bovaird & Loeffler, 2012; Davis & Ruddle, 2012; Leone et al., 2012; A. Meijer, 2013), and characteristics of participants including professionalism and trust in government (Brandsen et al., 2018; Lachmund, 1998; Schafft & Brown, 2000; Wise et al., 2012). However, what is relatively unknown is how organizational performance affects co-production. This is a critical area of studies that prior research has overlooked; past performance as a determinant of co-production in public organizations.

This study aims to investigate whether performance signals drive organizations to engage in co-production. Past performance is a critical signal for organizations due to managers' bounded rationality (Audia & Greve, 2021; Meier et al., 2015). Performance outcomes provide indicators for stakeholders including citizens and regulators to assess whether the activities of public organizations were satisfying their expectations. When the performance outcomes failed to meet the goals, public managers face accountability pressure and initiate problemistic search to explore strategies to improve future performance (Hong, 2019a; Nielsen & Baekgaard, 2015; Posen et al., 2018; van der Voet, 2022). In this setting, co-production, which is known to fit the current public management environments and positively influence future performance growth, becomes an appealing strategy in response to performance shortfalls.

This paper employs a regression discontinuity design (RDD) with New York City public school data to provide evidence for the relationship between performance shortfall and perception of co-production. The RD design is a quasi-experimental method to identify

how performance signals affect subsequent efforts for co-production. The results show that low-performing organizations are more likely to be willing to engage in the co-production.

This paper makes the following contributions to the literature on co-production and performance feedback. First, this article deepens our understanding of the drivers of co-production. Prior studies have paid little attention to past performance signals as a determinant of co-production. This paper finds that organizations show different behaviors on co-production depending on their performance. As willingness to co-produce is closely related to the success of co-production (Landi & Russo, 2022; Lindenmeier et al., 2021; O'Brien et al., 2016), it is important to tell which organizations have positive incentives toward co-production. Second, this paper contributes a comprehensive understanding of co-production through inclusion of citizen participation in the assessment. A growing body of scholarship demands a discussion of diverse actors in co-production; public officials and citizens are both important players in effective co-production (M. Jakobsen & Andersen, 2013; Voorberg et al., 2015). This study, hence, considers co-production based on both activities of public organizations and citizen participation. Third, by investigating how performance shortfalls affect co-production, this study also identifies a strategy that has been little known in response to performance shortfalls. Although performance management literature has examined several behaviors, such as re-prioritization of goals and innovation (Nicholson-Crotty et al., 2017; Nielsen, 2014a), co-production has not been highlighted as a strategy for low-performing organizations. It may also provide implications for low-performing organizations to improve their performance.

Literature Review on Co-production

Co-production is the process by which external stakeholders of public service actively involve in public service delivery with public sector professionals (Clark et al., 2013; M. Jakobsen & Andersen, 2013; Nederhand & Van Meerkerk, 2018; Ostrom, 1972; Pestoff, 2006; Voorberg et al., 2015). Although co-production seems to be similar to collaboration and citizen participation, co-production is a narrower concept compared to citizen participation and collaboration. First, co-production is goal-oriented; citizens or external actors engage in the co-production process to achieve a concrete goal with public employees (Gazley et al., 2020). It is differentiated from nonbinding collaboration such as the memorandum of understanding (MOU), which is a low level of collaboration. Second, co-production requires active roles for citizens (Nederhand & Van Meerkerk, 2018). In the co-production process, citizens are not limited to an advisory role. Instead, to improve public service quality, citizens participate in the whole management process such as accessing resources of public organizations, providing feedback to public organizations, finding solutions with regular producers, and taking a responsibility for the results (M. Jakobsen & Andersen, 2013). Public organizations, on the other hand, serve as supporter roles for co-production such as providing advice, information, and opportunity to learn about public service (Andersen et al., 2020; Bovaird, 2007; Edelenbos et al., 2017). Third, the meaning of citizens in co-production is close to the stakeholder who has strong interests on certain policies and public issues (Clark et al., 2013). It is different from an open-based collaboration such as an innovation community where anybody can propose problems and offer solutions. For example, when citizens and police officials co-produce the public

safety service, the citizens in this context are the stakeholder who is closely related to the safety policy and benefit from the policy.

While great strides have been made, current literature has limitations. In the literature, it is still unknown how past performance signals influence co-production. Although previous literature investigated the factors that contribute to initiating co-production, little research has been done on how past performance levels influence co-production. Previous literature on public management has highlighted several determinants of co-production. Those determinants can be broadly categorized into regular producer contexts and end-user contexts. First, the regular producer context includes organizational structure enabling co-production (Bovaird & Loeffler, 2012; A. J. Meijer et al., 2012), the attitude of public officials (Davis & Ruddle, 2012; Leone et al., 2012), social media usage (Cho & Melisa, 2021), and administrative culture (Baars, 2011). Public organizations are more likely to co-produce with citizens when they have an organizational structure for co-production, positive perceptions of co-production, and traditions of collaborating with citizens. Second, on the participants side, the factors include citizen characteristics, a sense of ownership, and trust in public organizations. Citizens are willing to participate in co-production when they have intrinsic characteristics such as a preference to engage in government decision-making and civic duty (Wise et al., 2012), responsibility for public service (Brandsen et al., 2018), and trust in the co-production system (Lachmund, 1998; Schafft & Brown, 2000).

This paper examines the effect of performance shortfalls in public organizations on co-production to address the gaps in the previous literature. Theoretically, performance shortfalls may work as a determinant of co-production. This is because co-production is a

plausible strategy for future performance growth in the current public management context. This study, thus, tests whether organizations seek to have a high level of organizational support and citizen participation in response to performance shortfalls. In what follows, this paper explains why low-performing organizations are likely to pursue co-production.

Performance feedback and Problemistic Search

Performance feedback, through which an organization receives information about the results of their organizational activities, is a significant signal that facilitates organizations to analyze their experiences (Greve, 2003). This is because past performance is an indicator of whether organizations have fulfilled goals within a specific timeframe; based on past performances, organizations judge whether their prior activities were appropriate or not (Choi et al., 2019). Public organizations acquire this information by measuring past activities quantitatively and qualitatively, and this information positively affects trust by obtaining confirmation from external actors such as regulators (Micheli and Mari 2014; Micheli and Pavlov 2020). One example is New York City school quality reports in which the New York City Department of Education assesses the performance of each school and assigns performance scores and grades, which indicate the levels of performance. Organizations, especially those that did not meet performance targets, consider past performance as a signal that their previous strategies and operations were problematic (Greve, 2003). This signal leads organizations to attempt to change their strategies (Meier et al., 2015; Nicholson-Crotty et al., 2017).

Organizations' recognition of performance shortfalls facilitates organizational learning; learning helps them find strategies to improve performance through searching behaviors (Cyert et al., 1963; Greve, 2003). Searching behaviors in low-performing organizations are problem-driven, which are called problemistic searches with the aim of discovering alternatives to overcome the problems (Posen et al., 2018). Problemistic search includes two types: problem-definition and solution-generation search. Problem definition is the process of identifying the reasons for the performance shortfall while solution generation is the process of finding the strategies that would restore the performance outcomes (van der Voet, 2022). Through problemistic search, low-performing organizations seek to fix these problems that may threaten managers' positions or authorities; they attempt to quickly recover their performance (Lehman & Ramanujam, 2009; Wang & Yeung, 2019).

In this setting, public administration studies should extend the examination on what solutions public organizations adopt after problemistic search. However, the discussion for the solution generation in response to the performance shortfall is mostly in debts to business literature. Previous literature on private sector management has verified diverse types of organizational changes from narrowly defined strategies (Arrfelt et al., 2013; Audia & Greve, 2006), to broadly defined changes, such as the structure and orientation of strategies (Audia et al., 2000). Examples of organizational changes in response to performance shortfalls include acquisition and mergers (Iyer & Miller, 2008; Kuusela et al., 2017), downsizing (Joseph et al., 2016), and increasing investment in R&D (Lucas et al., 2018). On the other hand, in previous public management literature, co-production has not been highlighted as a strategy for low-performing organizations. Instead, the previous

public management literature has investigated other forms of organizational behaviors in response to performance shortfalls, such as the re-establishment of an organization's goals (Nielsen, 2014a; Salge, 2011a) and organizational innovation (Nicholson-Crotty et al., 2017). Furthermore, prior studies on performance feedback have focused more on political factors and subsequent performance improvement. For instance, there is much evidence that public managers and citizens sensitively respond to performance shortfalls, which is called negativity bias (e.g. Holbein, 2016; Hong et al., 2020; Hong & Kim, 2019; Kogan et al., 2016; Lewandowski, 2019; van den Bekerom et al., 2021). This strand of research has also examined the direct relationship between performance feedback and subsequent performance (Hong, 2019a; Hong et al., 2020a; Min et al., 2021).

Although prior studies on business management suggested several types of solutions-generation behaviors in low-performing organizations, these solutions may not fit the public management context because public organizations have different characteristics compared to private organizations. Different from the business sector, a variety of stakeholders including regulators and citizens engage in public management, and this trait limits the discretion of public managers (Lyons et al., 2006). For example, it is relatively hard to lay employees off in public organizations, public organizations should receive permission from congress or supervisory organizations to change organization structure, and public organizations have constraints to attract investment from external actors (Selin & Lewis, 2018).

Co-production, Performance Outcomes, and Public Management Context

In the current public sector environment, co-production has several characteristics that make it an appealing strategy to respond to low-performing signals. First, low-performing organizations are often under significant accountability pressure. Struggling in effectively providing public services also hurt the image and legitimacy of these public organizations. Citizen participation has been proved to be an effectively approach to promote government trust, legitimacy, and management quality (Schmidthuber et al., 2021). Low-performing organizations are thus likely to invite citizens to participate in the processes of decision-making and delivery of public services in order to improve organizations' legitimacy and citizens' trust in these organizations.

Second, despite some inconsistent findings, co-production has been found to positively influence organizational performance, which is a useful remedy for low-performing organizations (Kang & Van Ryzin, 2019; Landi & Russo, 2022; Lindenmeier et al., 2021; Parrado et al., 2013). Through involving citizens in co-production process, public organizations gain better understanding of citizens' expectations, preferences and their knowledge, which can be instrumental in public service delivery (Dudau et al., 2019; Magnusson, 2003). It drives innovation in the public service and leads to better performance outcomes, and positively influences the performance outcomes since the expectations of end-users shape the quality of public service (Osborne et al., 2018).

Third, co-production fits the current public sector environment. The concept of co-production has a long tradition; scholars coined the term, co-production, to explain the role

of citizens in the process of public service provision in the 1970s and 1980s (Ostrom, 1972; Parks et al., 1981; Percy, 1978). This is because the economic conditions were not favorable for public organizations. Several economic shocks from the oil crisis and the Vietnam war required public organizations to constrain their budget and provide public service with limited resources (Alford, 1998). As a way to meet the demand for public service with limited resources, scholars considered inviting citizens into the process of producing public service. By the same token, the current environment around the public sector is similar to the era that the concept of co-production was coined. Economic and health crises, such as the 2008 subprime mortgage crisis and Covid-19, caused the cutback management again in the public sector (Maher et al., 2020; Nabatchi et al., 2017; Van der Voet, 2022). Co-production is thus an important strategy for public organizations in tough times.

Co-production, however, presents at a cost for public organizations as it produces adverse effects on internal management. First, co-production may require psychological costs for citizens; citizen participants in co-production may feel pressure from moral obligation and social norm to care about the benefits of their immediate acquaintances rather than the collective benefits for all community members (Thomsen et al., 2020). When participants weigh benefits of their acquaintances more than the whole community, it may distort effective decision-making in public organizations. Besides, participation that is not well organized or is structured with lower self-efficacy and feelings of empowerment also diminishes its effectiveness of co-production (Dias & Maynard-Moody, 2007; Svara, 2008). Second, co-production brings unintended impacts to public organizations. By sharing their information with participants, public organizations increases the possible

exposure to unjustified feedback from citizens and journalists (Grimmelikhuijsen, 2012). Sharing information with colleagues and the public can also induce employees to hide negative results and reveal only successful products (Bernstein, 2012). Overall, not all managers seek to co-produce with citizens because it does not guarantee a positive effect on their organizations. Based on the behavioral theory of the firm and performance feedback theory, low-performing organizations could engage in co-production to fix the urgent problem: performance improvement.

It is also necessary to consider how organizations respond to the high-performance signals. The behavioral theory of the firm indicates that organizations make a decision for their future behaviors based on the satisficing principle; when the results of behaviors are not satisfactory, organizations recognize them as a problem and engage in innovation for resolving them (Cyert et al., 1963; Greve, 2003). Therefore, high-performing and low-performing organizations may have different risk preferences when responding to their performance signals. Low-performing organizations are more likely to adopt a risk-tolerant solution to improve the performance outcomes (Hong, 2019a; Meier et al., 2015; Nicholson-Crotty et al., 2017). High-performing organizations, on the other hand, are interested in risk-averse options such as exploitation that refines existing strategies and (Greve, 2003; Gupta et al., 2006; D. Levinthal & March, 1981; March, 1991). In this sense, high-performing organizations may not recognize the performance information as the signal to initiate the novel policy. Rather, high-performing organizations have more interest in refining and making perfect on their ongoing policy (Lavie et al., 2010; March, 1991). Summarizing the above scenarios, this study proposes the following hypotheses:

Hypothesis 1: Public organizations increase the degree of co-production in response to poor performance signals

Hypothesis 2: Public organizations do not significantly change the degree of co-production in response to high performance signals

Backgrounds, Data, and Method

Parental involvement in public education is a classic case of co-production (Bifulco, 2006). Previous literature on public education has long emphasized parental involvement as an important element for successful school reform and school performance; public schools invite parents to improve school quality and student academic achievement (McMillan, 1999; Purkey & Smith, 1983). Parents of public schools are crucial stakeholders of public education because school management and student academic performance are closely related to their children (Andersen et al., 2020; M. Jakobsen, 2013). Moreover, parents have the knowledge and experience that traditional service providers who are teachers and principals in this case, may not have. Parents understand their children better, often help with students' homework and monitor their academic progress, which gives them unique knowledge about their children's academic potentials (Bifulco, 2006). Public schools, additionally, have incentives to actively support the parents for co-production. Previous literature showed that schools achieve positive outcomes from involving parents (Clark et al., 2013; Levine, 2018). Furthermore, co-production in education context helps staffs to develop a positive relationship with parents and learn from their heterogeneous experiences

and knowledge (Marschall, 2006). It is crucial for teachers and principals to gain confidence in their teaching and meet the special needs of their students (Gonzalez et al., 1995; Hoover-Dempsey et al., 1987). With the practical and theoretical backgrounds, this paper utilizes the public school case for co-production.

Empirical evidence for this paper comes from New York City school quality report, which includes the performance ratings and scores from 2014-15 AY to 2017-18 AY. New York City Department of Education (NYC DOE) annually issues this report online where citizens or interested parties are able to read it. In the frame of school quality reports from the 2014-15 AY to the 2017-18 AY, school performance includes four ratings, which include Exceeding Targets, Meeting Targets, Approaching Targets, and Not Meeting Targets, and ranges of scores were from 1.00 to 4.99. Ratings were based on performance scores, and each integer was a cutoff point for the rating. For instance, 2.00 is a cutoff point for not meeting target grade and approaching target grade. The subsequent two digits indicate the degree of performance within the performance rating; it shows how much the performance of schools is close to or far from the next grade. Suppose that school A scored 2.05 and school B scored 2.95. Although these two schools achieved the same grade, 'approaching target,' the performance of school B is likely to be similar to the higher grade, meeting the target, while that of school A is likely to not have much difference with a lower grade, not meeting the target.

The overall score of student achievement , which is the basis of ratings, is calculated with the weighted averages on multiple sources of achievement indicators and the additional credits that schools can receive when they reduce the achievement gaps between more advantaged and more disadvantaged groups (NYC Department of Education 2018).

The measurement for student achievement in elementary, K-8, and middle school includes proficiency rates in Math and English Language Arts, which indicate the percentages of students who achieve levels 3 and 4 in the tests, Average test scores in Math and ELA, and core course pass rates. In addition, schools can score additional credits when high-need students show progress in their tests; high-need students in this setting include minority groups in schools, such as English Language Learners, students with disabilities, and Black and Hispanic students. However, it is not substantial in the overall score where the maximum is 4.99; the available extra point is up to 0.25 points.

This paper utilizes student achievement performance in elementary, K-8, and middle schools. A major reason to exclude high schools is that high schools used different performance metrics such as graduation rates. Moreover, the sample of elementary, K-8, and middle schools allows this study to have a bigger sample size than the high school sample.

Though there is no formal sanctions or rewards associated with performance ratings, it is a form of public "naming and shaming" since the results get wide attention from the public and the media. Research has suggested that pressure from "naming and shaming" often incentivizes poor performance organizations to improve performance (Bohte & Meier, 2000a; Desmidt & Meyfroodt, 2021; Ladd & Lauen, 2010). The way to measure performance during the research periods was consistent. This study utilizes this dataset as one of the primary datasets to capture the changes at the cutoff points of performance rating.

Dependent variables

The dependent variables, co-production in this paper, come from New York City parent surveys from 2015-16 AY to 2018-19 AY. There is a one-year lag between dependent variables and independent variables because school performance ratings in year t are published in the fall semester of year t+1, which means the effects of performance ratings are likely to appear in year t+1. The NYCDOE annually surveys parents to gauge opinions on several topics, such as communication with their children's schools and satisfaction with school management and leadership.

This study uses four items to measure the degree of co-production, including two items for organization support to facilitate citizen participation and two items for the set of citizen participation activities. Table 1 presents the questions. The questions for organization support include the degree of accessibility to the classroom and the provision of information from the school. By accessing the workplace and receiving information about public services, citizens gain a deeper understanding of public services (Jakobsen, 2013; Jakobsen & Andersen, 2013). The items for participation consist of eliciting feedback from parents and communication about student performance. The first three questions ask participants to choose one that is closest to their opinion among four options: strongly disagree, disagree, agree, and strongly agree. The last question asks about the frequency, and the options include never, rarely, sometimes, and often. The empirical analysis in this study is based on positive rates, which indicate the sum of the percentages of "agree" and "strongly agree" for the first three questions and the sum of the percentage of "sometimes" and "often" in the item for the last question. The average responsive rate of parent surveys is approximately 62%. Although it is a slightly low value to generalize

the parents' thoughts, it is still the best measure we could acquire to gauge openness in schools.

Table 1-1 Survey Questions Related to Co-production

	Questions
Co-production	Q1. Parents/Guardians are invited to visit classrooms to observe instruction[1]
	Q2. My child's school will make me aware if there are any emotional or psychological issues affecting his/her academic performance.
	Q1. School staff regularly communicate with parents/guardians about how I can help students learn
	Q2. How often have you communicated with your child's teacher about your child's performance

Regression Discontinuity Design

The school quality report, which includes continuous performance scores and clear-cut ratings, enables this study to employ a Regression Discontinuity (RD) design. In this setting, RD estimates the differences in coproduction in schools whose performance scores are just above the cutoff or just below the cutoff. Although schools near the cutoff points achieve a similar level of performance, the performance ratings assigned by arbitrary rules send very different signals to principals, teachers, and parents. Schools, whose performance is just below the cutoff points, receive a lower rating and may be under greater accountability pressure, while schools whose performance is just above the cutoff points may be perceived

[1] This question was worded slightly differently in the 2018-19 AY. It reads 'My child's school offers me opportunities to visit my child's classroom, such as observing instruction, participating in an activity with my child, etc.'

as high-performing. We could thus attribute any differences in subsequent organizational changes to the performance signal. The RD design provides a credible approach to causally identifying how performance signals affect subsequent organizational changes.

The continuous performance score works as the running variable. This study employs sharp RD design because the scores to classify the ratings are consistent and clear-cut, and all observations comply with the treatment assignment (Cattaneo et al., 2019). The RD model in this study is a local polynomial regression that does not rely on parametric assumptions of the relationship between dependent variables and the running variable. It uses observations within a small bandwidth near the cutoff point to fit regressions (Cattaneo et al., 2019). Therefore, the local RD model utilizes observations around the cutoff point, enabling us to capture changes occurring near the cutoff point. One potential issue with this approach is that the results could be different depending on how bandwidths formulate the neighborhood of the cutoff point; it raises concerns about specification searching and ad-hoc decisions. For this reason, this study uses several methods for bandwidth selection outlined by Cattaneo (2019). Specifically, this study adopts three commonly-used bandwidths: Mean squared error (MSE)[2] optimal, MSE sum, and Coverage error (CER)[3] optimal. These bandwidths minimize the mean squared error, mean squared error of the sum of the regression coefficient, and asymptotic coverage error rate respectively. In addition, this study also adopts the local quadratic polynomial with MSE-

[2] Mean squared error occurs when there is an average squared difference between actual and estimated values.
[3] Coverage error occurs when the actual sample does not correspond to the target population.

optimal to test the robustness of the results. The local RD model employed in this study is as follows:

$$Y_{s(t+1)} = a + f(X_{s(t)}) + \beta G_{s(t)} + \gamma Covariates_t + \pi_t + \epsilon_{st}$$

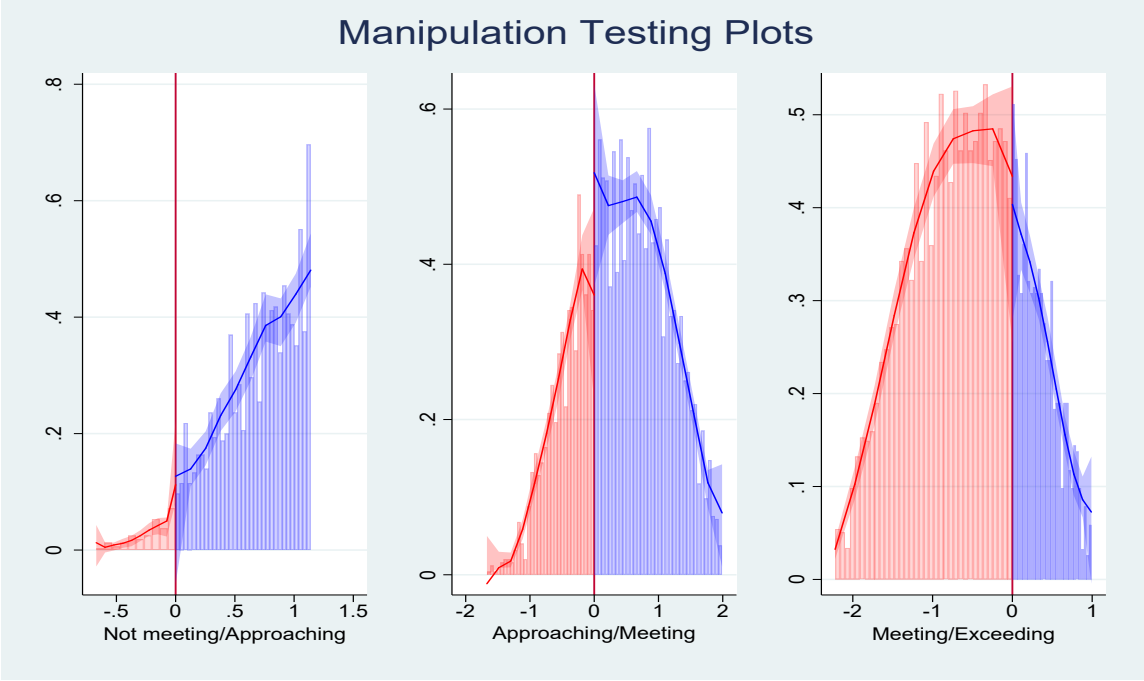
where $Y_{s(t+1)}$ indicates openness variables, which are positive rates of dependent variables in school s in year t+1; $f(X_{st})$ is a linear or quadratic polynomial of the running variable X, which is continuously distributed performance scores, in year t; and $G_{s(t)}$ is the grade determined by the running variable. The vector of covariates consists of school enrollment, the Economic Need Index, students' racial composition, which includes percentages of White, Black, Hispanic, and Asian students, English language learners (ELL), and students with a disability (SWD). This model also includes parent response rates as a covariate. π_s and π_t are the school and year fixed effects.

Validation Tests for Regression Discontinuity Design

The regression discontinuity design should satisfy two assumptions to validate the empirical model: no manipulation in the running variable and continuity of the regression function in the absence of the treatment. The first assumption is that units cannot manipulate their treatment status by changing their values of the running variable. If subjects can manipulate their performance scores below or above the cutoffs purposefully, it would violate this underlying assumption of the RD design. Furthermore, it raises doubts

about the interpretation of the results since it is hard to tell whether the results come from the discontinuity of dependent variables or the manipulation of running variables. The second assumption is the continuity of regression functions except for the treatment. There should be no abrupt changes in the regression functions at placebo cutoff points.

To test the first assumption, this paper employs the local polynomial density test at the cutoff points; the test is to check the differences in density at the cutoff points. If subjects had the ability to manipulate their performance scores, we should be able to detect significant jumps or drops at the cutoff points. Figure 1 and Table 2 present the results of the density test. Figure 1 shows that 95% confidence intervals overlap at the cutoff points, indicating that there were no abrupt changes at these points. These results support that subjects are less likely to manipulate their performance scores.



Note: Line and shading area indicate the point estimate and 95% confidence interval respectively

Figure 1-1 Density Plots at Cutoffs

Table 1-2 Manipulation Test Results

Cutoff	(1) Not meeting/Approaching	(2) Approaching/Meeting	(3) Meeting/Exceeding
T-value	-1.881	1.349	0.188
P-value	0.060	0.177	0.851
Effective N	375	3303	2672

Note: The null hypothesis is that there is no difference in the density of treated and control observations at the cutoff. Polynomial fit order of 4.

This study also employs placebo cutoffs near the actual cutoff points to demonstrate that there was no significant discontinuity in regression functions for treatment and controls except for the actual cutoff points. Specifically, this paper sets four placebo cutoffs – two points are below the actual cutoff and others are above the actual cutoff. This test follows the way of Cattaneo, Idrobo, and Titiunik (2019), which tests the placebos below the true cutoff only with control observations and also tests the placebos above the true cutoff only with treated observations. The placebo cutoffs are located at —0.05, -0.025, +0.025, and +0.05 away from the actual cutoff; it sets performance scores, for example, 1.95, 1.975, 2.025, and 2.05, as the placebo cutoffs of the rating between not meeting target and approaching target grade. Table 3 presents the RD results with placebo cutoffs and they did not show a significant discontinuity at the 95% confidence interval.

Table 1-3 Results with placebo cutoffs

	(1) Parents' Accessibility to Classroom	(2) Provision of Information	(3) Feedback from Parents	(4) Communication on Students' Performance
Cutoff at not meeting and approaching target				
Alternative	-.092	3.176	.066	1.771
cutoff : **1.95**	[-11.55, 4.33]	[-4.36, 8.97]	[-6.4, 3.44]	[-2.62, 5.22]
Effective N	37	37	37	33
Alternative	3.557	.269	1.713	-.003
cutoff : **1.975**	[-4.4, 11.28]	[-4.49, 5.52]	[-3.24, 5.98]	[-3.64, 2.97]
Effective N	58	46	50	50
Alternative	4.751	3.971	5.43	2.878
cutoff : **2.025**	[-12.53, 22.76]	[-7.27, 15.04]	[-5.39, 19.24]	[-6.43, 14.67]
Effective N	1106	664	1406	776
Alternative	2.557	-.288	-.085	.803
cutoff : **2.05**	[-9.81, 14.39]	[-15.77, 3.67]	[-10.3, 4.52]	[-8.3, 5.93]
Effective N	189	221	228	189
Cutoff at approaching and meeting target				
Alternative	-.012	-.197	.422	.026
cutoff : **2.95**	[-10.3, 3.94]	[-7.93, 3.29]	[-3.62, 4.31]	[-2.79, 5.19]
Effective N	379	379	379	348
Alternative	4.234	.862	.272	-.901
cutoff : **2.975**	[-5.98, 5.23]	[-5.08, 4.54]	[-2.55, 5.81]	[-8.13, .87]
Effective N	313	331	331	331
Alternative	-.233	.067	.83	1.862
cutoff : **3.025**	[-5.61, 8.52]	[-2.59, 5.75]	[-3.19, 6.45]	[-3.71, 11.17]
Effective N	583	541	609	609
Alternative	-.364	1.564	.917	1.64
cutoff : **3.05**	[-6.86, 4.29]	[-1.77, 6.46]	[-4.29, 3.53]	[-4.73, 5.3]
Effective N	541	541	556	541
Cutoff at meeting and exceeding target				
Alternative	5.346	2.156*	.748	.717
cutoff : **3.95**	[-5.59, 11.01]	[-.22, 7.89]	[-5.37, 2.42]	[-5.21, 4.81]
Effective N	610	591	539	539
Alternative	3.579	1.611	2.045*	.539
cutoff : **3.975**	[-2.07,13.28]	[-1.54, 5.71]	[0, 7.08]	[-4.73, 5.47]
Effective N	591	659	568	591
Alternative	-4.114	-1.756	-1.437	-3.135
cutoff : **4.025**	[-12.19, 7.35]	[-5.49, .56]	[-4.43, 2.92]	[-4.87, 3.47]
Effective N	299	299	299	286
Alternative	-2.108	.775	.081	.794*
cutoff : **4.05**	[-7.09, 3.77]	[-1.03, 6.67]	[-1.83, 5.87]	[-.01, 11.63]
Effective N	320	320	320	299

Results

Descriptive Statistics

Table 4 shows the descriptive statistics of dependent variables and covariates. For control variables, schools, on average, have approximately 570 students and consist of 12%, 31%, 41%, and 13% of Asian, African American, Hispanic, and White students. Schools on average have approximately 14% and 22% of students categorized as English language learners and students with disability. The economic need index, which estimates the percentage of students who face economic hardship at a school, is about 0.68 during the research period. Approximately 62% of parents participated in the survey, which seems to have a low response rate, so this paper conducts alternative analyses via robustness checks, which include RD models with teacher survey items on coproduction, with parent response rates as the dependent variable to check if parent response rates were balanced at cutoff points, and without schools having low parent response rate. The table also shows that the dependent variables and covariates were stable over the research period. This means that there were less likely to be external events that affected school management in NYC public schools.

Table 1-4 Descriptive Statistics

		2015-16 AY (N = 1,199)	2016-17 AY (N = 1,206)	2017-18 AY (N=1,190)	2018-19 AY (N=1,180)
Dependent variable					
Access to Classroom	Mean	80.11 (10.75)	80.28 (10.75)	87.66 (7.97)	87.98 (7.88)
Provision of Information	Mean	84.74 (5.21)	88.11 (4.75)	88.01 (5.27)	88.35 (5.13)
Feedback from Parents	Mean	90.83 (4.59)	91.06 (4.68)	91.08 (4.95)	91.35 (4.80)
Communication on Student Performance	Mean	91.61 (4.81)	91.16 (5.33)	91.28 (5.12)	92.10 (5.47)
Control variable					
Survey response rate (%)	Mean	59.43 (21.62)	61.16 (21.49)	62.29 (21.60)	62.34 (21.94)
Enrollment	Mean	618.80 (343.39)	567.12 (343.75)	567.30 (344.49)	561.62 (338.22)
Economic need index	Mean	0.67 (0.21)	0.66 (0.21)	0.64 (0.22)	0.74 (0.22)
Asian students (%)	Mean	0.12 (0.18)	0.12 (0.18)	0.12 (0.18)	0.12 (0.18)
African American student(%)	Mean	0.33 (0.29)	0.32 (0.29)	0.31 (0.28)	0.30 (0.28)
Hispanic student (%)	Mean	0.41 (0.26)	0.41 (0.26)	0.42 (0.26)	0.42 (0.26)
White student (%)	Mean	0.13 (0.20)	0.13 (0.20)	0.13 (0.20)	0.14 (0.20)
English language learner (%)	Mean	0.13 (0.12)	0.13 (0.12)	0.14 (0.12)	0.14 (0.12)
Students with disability (%)	Mean	0.21 (0.07)	0.21 (0.07)	0.22 (0.07)	0.22 (0.07)

Figure 2 presents the distribution of performance scores and the ratings for each year. The distributions show similar patterns over the years. In the 2014-15 AY, about 1%, 25%, 53%, and 21% of public schools were assigned not meeting, approaching,

meeting, and exceeding target grades. The percentages of each rating were 1%, 21%, 46%, and 32% in the 2015-16 AY. In the 2016-17 AY, approximately 3%, 32%, 45%, and 20% of public schools respectively received grades from not meeting the target to exceeding the target, and the percentages of each rating were 3%, 30%, 45%, and 22% in the 2017-18 AY.

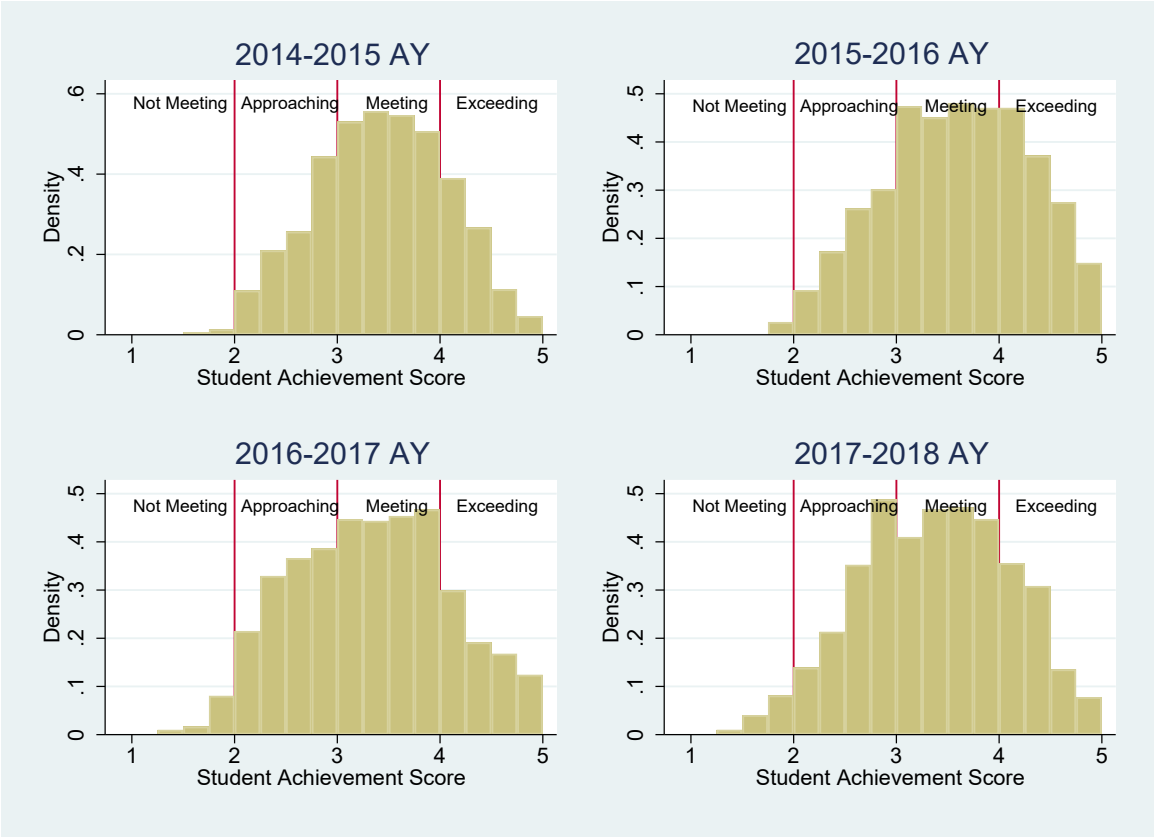


Figure 2 Distribution of Student Achievement Scores and Grades

RD Results for Organization Supports

The RD results presented in Table 5 indicate that poor-performing schools have a higher level of co-production in the form of organization support than average-performing schools. The degree of organization support significantly dropped at the not meeting/approaching cutoff. More specifically, regarding accessibility to the classroom, the positive rates in schools with "not meeting target" grade (the relatively lower rating at this cutoff) are higher than the scores in schools with "approaching target" grade (the relatively higher rating) by 5.4 percentage points (pp) to 7.2pp. When it comes to the provision of information, schools with "not meeting target" rating had higher positive rates compared to schools with "approaching target" rating by 3.8pp to 4.4pp. The evidence shows that poor-performing schools are more likely to open their classroom to parents and provide information about their children. In other words, parents whose children are in a poor-performing school are more likely to agree that they can access the classroom and that schools inform them about their child's school life. These results are consistent across different bandwidths and the polynomial functional form. However, at the other two cutoff points that determine approaching/meeting and meeting/exceeding target ratings, the RD model did not detect significant discontinuities, which means that there were no differences in the degree of coproduction in terms of organization support between schools that received "approaching targe" and "meeting target" ratings and between schools received "meeting target" and "exceeding target" ratings. In sum, this suggests that schools receiving "not meeting target" grade, which is the lowest performance rating, may expand their degree of the organization support to parents who are external actors of public schools, while schools receiving higher grades are less likely to do so.

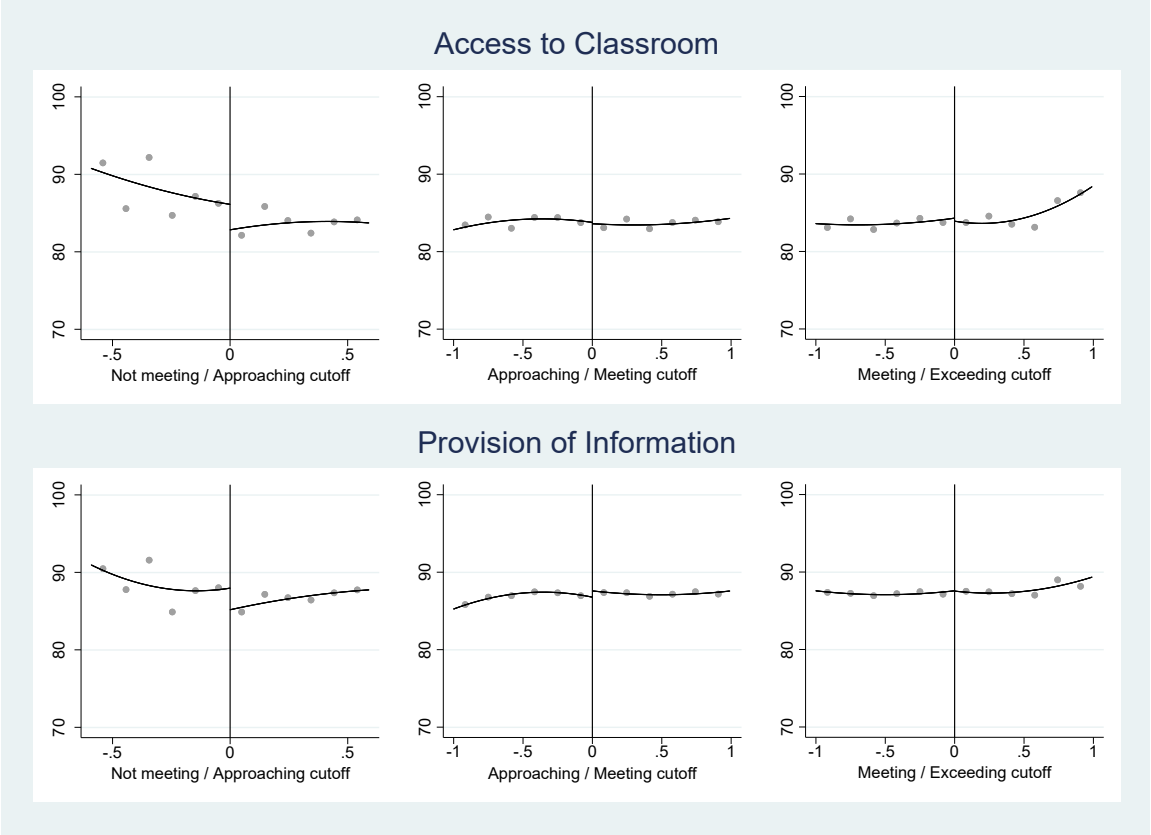
Table 5 RD Regression Results for Organization Supports

Dependent variables	(1) MSE-optimal bandwidth (mserd) Linear	(2) MSE-optimal bandwidth (msesum) Linear	(3) CER-optimal bandwidth (cerrd) Linear	(4) MSE-optimal bandwidth (mserd) Quadratic
Access to Classroom				
RD estimator	-5.37***	-3.945**	-6.377***	-7.181***
(Not meeting/Approaching)	[-10.84, -1.49]	[-9.49, -.6]	[-11.83, -1.7]	[-14.83, -2.09]
Effective N	191	234	133	234
RD estimator	.122	.134	-.246	-.426
(Approaching/Meeting)	[-1.77, 1.94]	[-1.66, 1.95]	[-2.24, 1.73]	[-3.09, 1.68]
Effective N	1771	1833	1213	1898
RD estimator	-.52	-.505	-.461	-.936
(Meeting/Exceeding)	[-2.88, 1.86]	[-2.86, 1.83]	[-3.05, 2.15]	[-4.72, 2.28]
Effective N	1591	1631	1146	1439
Provision of Information				
RD estimator	-4.4***	-3.779***	-4.711***	-4.615**
(Not meeting/Approaching)	[-8.63, -2]	[-7.77, -1.43]	[-8.84, -1.73]	[-9.24, -1.06]
Effective N	156	191	112	250
RD estimator	.556	.534	.525	.553
(Approaching/Meeting)	[-.65, 1.83]	[-.65, 1.78]	[-.81, 1.89]	[-.98, 1.98]
Effective N	1571	1650	1112	2119
RD estimator	.429	.432	.4	.339
(Meeting/Exceeding)	[-.56, 1.56]	[-.56, 1.56]	[-.7, 1.57]	[-1.08, 1.6]
Effective N	1669	1591	1176	1848

Note: Inference based on robust bias-corrected confidence intervals that are presented in brackets. All models include year fixed effects and covariates. Standard errors clustered at the school level. * p < 0.1 , ** p < 0.05, *** p < 0.01

Figure 3 shows the RD plots that depict the discontinuities with graphs at each cutoff point. Compared to the graphs for approaching/meeting cutoff and meeting/exceeding cutoff, the graph for not meeting/approaching cutoff shows a relatively large gap at the cutoff point. In other words, higher percentages of parents had access to their child's classroom or received information related to their child's school life in poor-performing schools than in average-performing schools. The bins, which are the local

sample mean in the graphs, seem to continue over the cutoff points that distinguished approaching/meeting ratings and meeting/exceeding target ratings.



Note: The dots indicate local averages of positive rates. Polynomial fit order of 2

Figure 3 RD plots for organization supports

RD Results for Participation

The RD results for coproduction in terms of participation are presented in Table 6. Parents whose children were in a poor-performing school were more likely to agree that schools regularly elicited feedback from them and frequently communicated with them. Specifically, with regard to the item gauging feedback from parents, the positive rates in

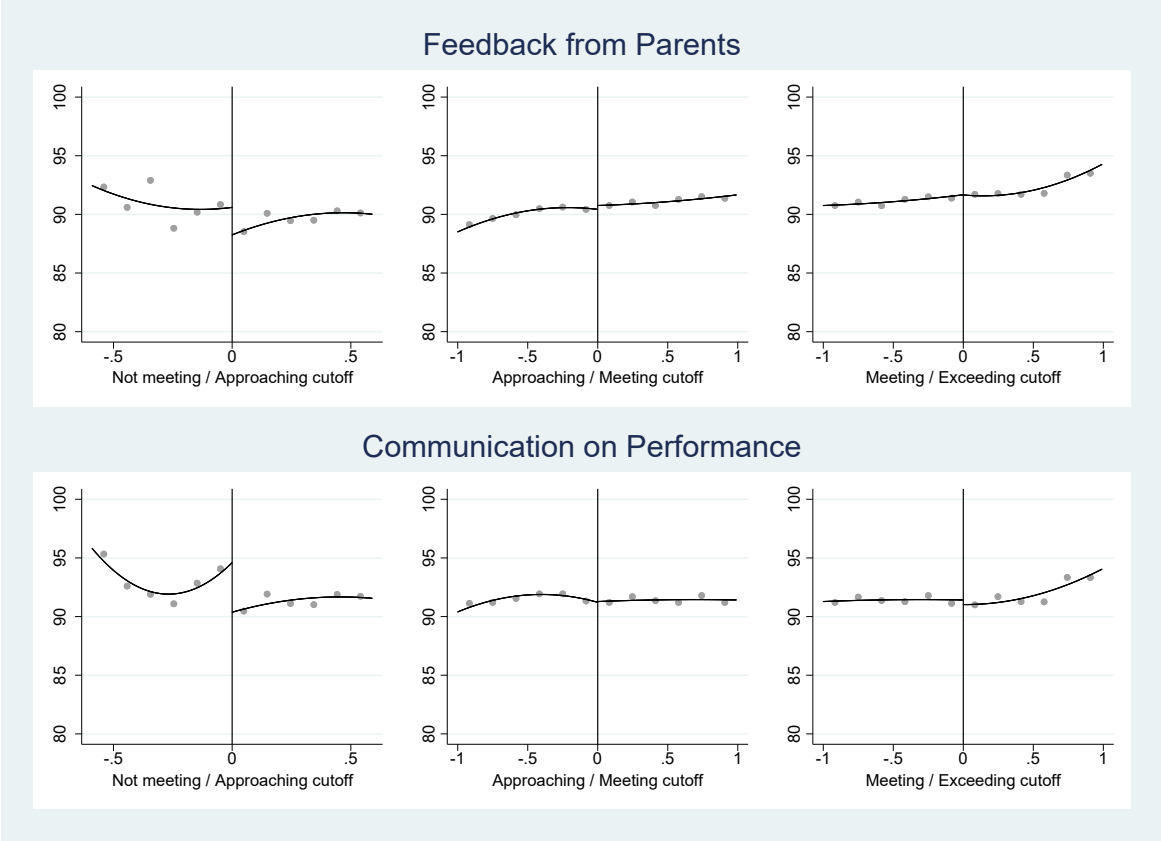
schools with “not meeting target” grade are higher than the positive rates in schools with “approaching target” grade by 4.1pp to 6.0pp. The RD results with the variable measuring the frequency of communication also show higher positive rates in schools with “not meeting target” rating compared to schools with “approaching target” rating by 3.7pp to 5.1pp. However, similar to the RD results for organization support, there were no significant discontinuities at the other cutoff points, which means the degrees of participation do not significantly change at the approaching/meeting target cutoff and meeting/exceeding target cutoff. In sum, schools that were assigned “not meeting target” grade may encourage parents’ participation such that they ask parents to provide feedback about student learning in regular meetings and frequently communicate about the child’s academic performance. The results are consistent across three different bandwiths and the quadratic functional form.

Table 6 RD Regression Results for Participation

Dependent variables	(1) MSE-optimal bandwidth (mserd) Linear	(2) MSE-optimal bandwidth (msesum) Linear	(3) CER-optimal bandwidth (cerrd) Linear	(4) MSE-optimal bandwidth (mserd) Quadratic
Feedback from Parents				
RD estimator	-4.594***	-4.127***	-5.88***	-5.969***
(Not meeting/Approaching)	[-8.34, -1.99]	[-7.75, -1.69]	[-9.52, -2.88]	[-10.89, -2.42]
Effective N	156	191	112	195
RD estimator	.445	.426	.507	.488
(Approaching/Meeting)	[-.57, 1.59]	[-.51, 1.54]	[-.62, 1.69]	[-.79, 1.73]
Effective N	1571	1725	1112	2198
RD estimator	.039	.038	-.205	-.432
(Meeting/Exceeding)	[-1.09, 1.12]	[-1.08, 1.12]	[-1.4, .97]	[-2.08, .85]
Effective N	1399	1399	995	1399
Communication on Performance				
RD estimator	-3.726***	-4.243***	-4.485***	-5.057***
(Not meeting/Approaching)	[-6.75, -1.44]	[-7.3, -1.84]	[-7.51, -1.82]	[-9.4, -1.89]
Effective N	195	180	133	216
RD estimator	.148	.139	.003	-.032
(Approaching/Meeting)	[-.9, 1.34]	[-.86, 1.35]	[-1.17, 1.26]	[-1.65, 1.26]
Effective N	1898	1940	1332	1972
RD estimator	-.548	-.455	-1.034	-.742
(Meeting/Exceeding)	[-2.36, .85]	[-2.31, .89]	[-2.87, .62]	[-2.69, .83]
Effective N	1109	1176	752	1704

Note: Inference based on robust bias-corrected confidence intervals that are presented in brackets. All models include year fixed effects and covariates. Standard errors clustered at the school level. * p < 0.1 , ** p < 0.05, *** p < 0.01

As Figure 4 presents, the results show relatively large gaps at the not meeting/approaching cutoff compared to gaps at other cutoff points. Specifically, the degrees of parents' feedback and communication with schools seem to be higher in schools receiving "not meeting target" grades than schools receiving "approaching target" grades. On the other hand, the degrees of parent feedback and communication seem to continue over other grade cutoffs smoothly.



Note: The dots indicate local averages of positive rates. Polynomial fit order of 2

Figure 4 RD plots for participation

Robustness Check

Testing the effects of survey response rates

This study also conducts several tests to make sure that uneven response rates did not cause bias in RD results. Although the response rates have been relatively consistent over the years, which is approximately 62%, the range of the response rate varies from school to school, from 1% to 100%. First, this paper runs RD regressions without schools with response rates below 30% to mitigate these threats. Admittedly, 30% is an arbitrary threshold, but we deem that response rates below 30% are more unreliable. Table 8 presents

the results with the samples excluding the very low response rates. The results are consistent with the results in the main RD model.

Furthermore, the results from the main model may be biased if the response rates are not balanced at each cutoff point. Hence, this study estimates RD regressions with parent survey response rate as the dependent variable to check the balance of this variable at the cutoff points. If the response rates do not significantly change at the cutoff points, the rates are balanced. Results presented in Table 9 support that parent response rates are balanced at the cutoff points.

Table 8 RD Regression Results without schools with very low response rates (< 30%)

Dependent variables	(1) MSE-optimal bandwidth (mserd) Linear	(2) MSE-optimal bandwidth (msesum) Linear	(3) CER-optimal bandwidth (cerrd) Linear	(4) MSE-optimal bandwidth (mserd) Quadratic
Access to Classroom				
RD estimator	-5.925***	-5.813***	-6.899***	-6.331**
(Not meeting/Approaching)	[-11.86, -1.9]	[-11.46, -1.88]	[-12.78, -2]	[-12.93, -1.22]
Effective N	151	166	117	241
RD estimator	.017	.019	-.431	-.661
(Approaching/Meeting)	[-1.98, 1.83]	[-1.98, 1.83]	[-2.52, 1.58]	[-3.45, 1.53]
Effective N	1656	1656	1158	1741
RD estimator	-.332	-.332	-.12	-.429
(Meeting/Exceeding)	[-2.71, 2.04]	[-2.71, 2.04]	[-2.73, 2.5]	[-3.92, 2.87]
Effective N	1552	1552	1084	1552
Provision of Information				
RD estimator	-4.25***	-3.973***	-4.321**	-4.279*
(Not meeting/Approaching)	[-8.43, -1.65]	[-8.12, -1.49]	[-8.46, -1.15]	[-8.95, .38]
Effective N	151	176	106	185
RD estimator	.75	.736	.633	.72
(Approaching/Meeting)	[-.52, 2.12]	[-.52, 2.08]	[-.79, 2.1]	[-.82, 2.18]
Effective N	1421	1451	984	2043
RD estimator	.443	.5	.26	.23

(Meeting/Exceeding)	[-.68, 1.69]	[-.68, 1.76]	[-.96, 1.54]	[-1.46, 1.54]
Effective N	1271	1114	865	1332
Feedback from Parents				
RD estimator	-4.361***	-4.378***	-5.521***	-5.09**
(Not meeting/Approaching)	[-8.21, -1.64]	[-8.14, -1.65]	[-9.28, -2.41]	[-9.82, -1.14]
Effective N	146	146	106	185
RD estimator	.321	.321	.273	.296
(Approaching/Meeting)	[-.8, 1.43]	[-.79, 1.43]	[-.93, 1.46]	[-1.06, 1.51]
Effective N	1451	1451	1026	2043
RD estimator	-.041	-.041	-.133	-.291
(Meeting/Exceeding)	[-1.05, 1]	[-1.04, 1]	[-1.24, .99]	[-1.81, .94]
Effective N	1474	1474	1050	1552
Communication on Performance				
RD estimator	-3.212**	-3.706***	-3.766**	-4.06**
(Not meeting/Approaching)	[-6.2, -.78]	[-6.86, -1.09]	[-6.86, -.94]	[-8.29, -.7]
Effective N	179	146	125	202
RD estimator	-.006	.005	-.277	-.508
(Approaching/Meeting)	[-1.25, 1.14]	[-1.07, 1.19]	[-1.59, .99]	[-2.25, .8]
Effective N	1588	1805	1119	1683
RD estimator	-.492	-.438	-.885	-.559
(Meeting/Exceeding)	[-2.31, .93]	[-2.27, .95]	[-2.74, .79]	[-2.41, 1.07]
Effective N	1084	1151	745	1749

Note: Inference based on robust bias-corrected confidence intervals that are presented in brackets. All models include year fixed effects and covariates. Standard errors clustered at the school level. * p < 0.1 , ** p < 0.05, *** p < 0.01

Table 9 Testing the continuity of survey response rates at cutoffs

Dependent variables	(1) MSE-optimal bandwidth (mserd) Linear	(2) MSE-optimal bandwidth (msesum) Linear	(3) CER-optimal bandwidth (cerrd) Linear	(4) MSE-optimal bandwidth (mserd) Quadratic
Parent Response Rates				
RD estimator	.632	.847	-.604	.461
(Not meeting/Approaching)	[-9.43, 12.45]	[-8.87, 12.66]	[-11.62, 10.81]	[-15.91, 16.72]
Effective N	234	250	164	243
RD estimator	-.691	-.977	-2.11	-3.286
(Approaching/Meeting)	[-5.56, 4.21]	[-6.22, 4.06]	[-7.45, 3.23]	[-11.75, 3.76]
Effective N	2198	2010	1571	1802
RD estimator	-.19	-.227	-.423	-1.102
(Meeting/Exceeding)	[-7.18, 5.88]	[-7.27, 5.69]	[-7.73, 6.46]	[-8.39, 4.58]
Effective N	945	915	679	1704

Note: Inference based on robust bias-corrected confidence intervals that are presented in brackets. All models include year fixed effects and covariates. Standard errors clustered at the school level. * p < 0.1 , ** p < 0.05, *** p < 0.01

RD model with alternative dependent variables

Finally, responses to surveys are subjective and parents' responses may not represent the full extent of schools' efforts in coproduction. To triangulate the results, this paper utilizes surveys of teacher as another measure of school coproduction. If results from using similar questions on participation and organization supports in the teacher survey converge with the results from parent survey, we can be more confident about the results based on parent sruvey. The following table indicates the survey items in the teacher survey, which are similar to the items in the parent survey. In Table 10, question 1 is to measure the concept of organization support, and question 2 is about the concept of participation. The RD model with alternative variables have the same specification as the main model. The results, presented in Table 11, show that teachers in schools receiving "not meeting target" grade are more likely to agree that parents can access their classroom and that their school asks

for feedback via regular meetings. There are no significant differences at other cutoff points. The results are consistent with the results based on survey items from parent survey.

Table 10 Questions in Teacher Survey Related to Governmental Openness

	Questions in Teacher Survey
Governmental Openness	Q1. parents/guardians are invited to visit classrooms to observe the instructional program.[4]
	Q2. The principal/school leaders encourage feedback through regular meetings with parent and teacher leaders

[4] This question was worded slightly differently in the 2018-19 AY. It reads 'Parents/guardians are offered opportunities to visit their child's classroom, such as observing instruction, participating in an activity with their child, etc.'

Table 11 RD regression results with Teacher Survey items

Dependent variables	(1) MSE-optimal bandwidth (mserd) Linear	(2) MSE-optimal bandwidth (msesum) Linear	(3) CER-optimal bandwidth (cerrd) Linear	(4) MSE-optimal bandwidth (mserd) Quadratic
Access to Classroom				
RD estimator	-8.67**	-6.274*	-14.221***	-13.542**
(Not meeting/Approaching)	[-20.95, -1.4]	[-18.16, .5]	[-25.35, -6.23]	[-29.2, -3.54]
Effective N	142	164	100	190
RD estimator	.772	.766	.211	-.456
(Approaching/Meeting)	[-2.53, 4.1]	[-2.06, 4.29]	[-3.26, 3.68]	[-5.46, 3.09]
Effective N	1721	1862	1212	1568
RD estimator	.296	.208	.074	.598
(Meeting/Exceeding)	[-3.04, 4.26]	[-3.16, 4.04]	[-3.53, 3.99]	[-3.69, 4.56]
Effective N	1107	1144	750	1397
Feedback from Parents and Teacher Leaders				
RD estimator	-13.277**	-9.248**	-24.185***	-17.973**
(Not meeting/Approaching)	[-31.14, 2.94]	[-25.48, -.16]	[-43.2, -9.08]	[-40.6, -4.34]
Effective N	112	147	89	180
RD estimator	1.913	1.896	2.302	2.342
(Approaching/Meeting)	[-.89, 5.22]	[-.67, 5.23]	[-.86, 5.74]	[-1.55, 6.22]
Effective N	1767	1829	1254	2038
RD estimator	1.625*	1.633*	1.1	2.072*
(Meeting/Exceeding)	[-.12, 4.33]	[-.1, 4.46]	[-1.07, 3.73]	[-.4, 4.78]
Effective N	1290	1290	913	1874

Note: Inference based on robust bias-corrected confidence intervals that are presented in brackets. All models include year fixed effects and covariates. Standard errors clustered at the school level. * p < 0.1 , ** p < 0.05, *** p < 0.01

Discussion and Conclusion

RD results suggest that public organizations change their level of engaging citizens in co-production in light of past performance signals. Specifically, schools that received a poor performance rating showed a higher degree of co-production measured by organization support and citizen participation than schools with relatively higher performance rating one year after receiving performance ratings. They are more likely to give access to the

classroom to parents, provide parents with information on their children's performance and elicit feedback from parents. The results provide evidence that low-performing organizations may strategize their use of coproduction. In contrast, schools with high performance ratings, such as "exceeding target", "meeting target", and "approaching target" did not show differences in terms of coproduction efforts.

The findings suggest that involving citizens in co-producing public services can be an attractive strategy for low-performance organizations. As we discussed previously, coproduction can improve transparency and legitimacy of struggling public organizations; moreover, it may also help improve organizational performance by better understanding citizens' preferences and utilizing their information and feedback. This study thus contributes to performance feedback literature by identifying a novel strategy in response to performance feedback. Some critics have argued that considering organizational performance as the ultimate dependent variable causes biases when estimating the effects of management on performance output. A bourgeoning body of scholarship has focused on how past performance signals affect public management (Meier et al., 2015; Nicholson-Crotty et al., 2017; van der Voet, 2022). Nevertheless, co-production has not been highlighted in the literature on performance feedback. Instead, previous literature has been more interested in examining the final outputs, such as subsequent performance and political outputs (e.g. Holbein, 2016; Hong, 2019). In this sense, this study fills a gap by examining the hidden process in the path from performance feedback to subsequent outputs.

These findings deepen our understanding of the determinant of co-production. Different from prior studies that focused more on the effects of attitudes of public officials and citizens on co-production, this paper examined how past performance signals affect

the degree of co-production. The results show that the degree of co-production varies by the level of past performance. Furthermore, this study established a theoretical framework between performance management literature and co-production literature. Based on findings in performance management literature that the problemistic search in low-performing organizations (Greve, 2003; Hong, 2019a; Hong et al., 2020a; Nielsen & Baekgaard, 2015), this study argues that low-performing organizations may consider co-production as a strategy to overcome the critical moment and improve the performance.

An interesting point in the results is that schools were not responsive at the cutoff points that indicate high performance. Recent studies indicate that high-performing organizations also innovate to keep their current performance level or achieve more excellent outputs (e.g. Nicholson-Crotty et al., 2017). It raises the question on why high-performing organizations did not show significant changes at the cutoff points. It may be related to the different behavioral patterns between low- and high-performing organizations. Low-performing organizations attempt to explore alternatives to reform their routines because of their experience of failure (Argote & Greve, 2007; Meier et al., 2015; Miner et al., 1996), while high-performing organizations focus on refining their existing routines (Baum & Dahlin, 2007; D. A. Levinthal & March, 1993). That is, the innovation pattern in high-performing organizations focuses on making their routines perfect, such as reducing costs and keeping a long-term durable growth path (Birhanu et al., 2016; O'Brien & David, 2014). In this setting, co-production may not be an attractive strategy for decision-makers. Promoting organization support and participation commonly increases the number of actors who influence decision-making. As mentioned earlier, co-production brings risks in that external actors could present unjustified feedback and distort

the decision-making process (Grimmelikhuijsen, 2012; Svara, 2008). The risks that come from involving external actors may lead high-performing organizations to search for other moderate ways to refine their routines.

The evidence that low-performing organizations are willing to engage in co-production gives us another policy implications. Government needs to design the program that supports low-performing organizations to initiate and manage the co-production process. It would nurture democratic values, enhance self-efficacy of citizens, and also restore performance outcomes of underperforming organizations.

Despite these contributions, this study has limitations. A notable one is that dependent variables are based on a perceptual measurement. Compared to archival measurement, perceptual data, including surveys and interviews, are likely to be biased by participants' backgrounds. Although this study uses multiple survey items to minimize the biases from perceptual data, the fact that the analyses are based only on parent surveys may threaten the validity. However, it does not mean that perceptual data are of lower quality than archival data. Perceptual measurement complements archival measurement in that it allows us to measure the subjective level of individuals, which is difficult to determine with archival data (Meier & O'Toole, 2013). Furthermore, co-production has been thought of as a concept that is hard to estimate with archival measurements. Previous literature on co-production reflects these hardships. To measure the concept of co-production such as citizen and public official attitudes, many prior studies relied on perceptual variables (Landi & Russo, 2022; Lindenmeier et al., 2021; O'Brien et al., 2016; Thomsen & Jensen, 2020). For instance, the level of accessibility to classrooms and the level of communication between parents and teachers are concepts that are hard to measure using archival datasets.

Finally, to allieviate concerns about statistical power, this paper set several alternative models, including validation tests and robustness checks to support the results of the main RD model. Although these attempts and explanations may not be sufficient to eliminate concerns about the validity of this paper's empirical model, this paper is expected to be of greater use if the evidence is supported by future studies.

Reference

Adler, P. S., & Borys, B. (1996). Two types of bureaucracy: Enabling and coercive. *Administrative Science Quarterly*, 61–89.

Alford, J. (1998). A public management road less travelled: Clients as co-producers of public services. *Australian Journal of Public Administration*, *57*(4), 128–137.

Andersen, S. C., Nielsen, H. S., & Thomsen, M. K. (2020). How to increase citizen coproduction: Replication and extension of existing research. *International Public Management Journal*, *23*(5), 696–712. https://doi.org/10.1080/10967494.2018.1518851

Andersson, T., & Liff, R. (2012). Multiprofessional Cooperation and Accountability Pressures: Consequences of a post-new public management concept in a new public management context. *Public Management Review*, *14*(6), 835–855. https://doi.org/10.1080/14719037.2011.650053

Argote, L., & Greve, H. R. (2007). *A Behavioral Theory of the Firm* —40 Years and Counting: Introduction and Impact. *Organization Science*, *18*(3), 337–349. https://doi.org/10.1287/orsc.1070.0280

Arrfelt, M., Wiseman, R. M., & Hult, G. T. M. (2013). Looking backward instead of forward: Aspiration-driven influences on the efficiency of the capital allocation process. *Academy of Management Journal*, *56*(4), 1081–1103.

Audia, P. G., & Greve, H. R. (2006). Less likely to fail: Low performance, firm size, and factory expansion in the shipbuilding industry. *Management Science*, *52*(1), 83–94.

Audia, P. G., & Greve, H. R. (2021). *Organizational Learning from Performance Feedback: A Behavioral Perspective on Multiple Goals: A Multiple Goals Perspective* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781108344289

Audia, P. G., Locke, E. A., & Smith, K. G. (2000). The paradox of success: An archival and a laboratory study of strategic persistence following radical environmental change. *Academy of Management Journal*, *43*(5), 837–853.

Baars, T. (2011). Experiential science; towards an integration of implicit and reflected practitioner-expert knowledge in the scientific development of organic farming. *Journal of Agricultural and Environmental Ethics*, *24*(6), 601–628.

Bajorek, Z. M., & Bevan, S. M. (2015). Performance-related-pay in the UK public sector: A review of the recent evidence on effectiveness and value for money. *Journal of Organizational Effectiveness: People and Performance*.

Barnow, B. S., & Heinrich, C. J. (2010). One Standard Fits All? The Pros and Cons of Performance Standard Adjustments. *Public Administration Review*, *70*(1), 60–71. https://doi.org/10.1111/j.1540-6210.2009.02111.x

Baum, J. A. C., & Dahlin, K. B. (2007). Aspiration Performance and Railroads' Patterns of Learning from Train Wrecks and Crashes. *Organization Science*, 19.

Belle, N., & Cantarelli, P. (2017). What Causes Unethical Behavior? A Meta-Analysis to Set an Agenda for Public Administration Research. *Public Administration Review*, *77*(3), 327–339. https://doi.org/10.1111/puar.12714

Benaine, S. L., & Kroll, A. (2020). Explaining effort substitution in performance systems: The role of task demands and mission orientation. *Public Management Review*, *22*(6), 813–835. https://doi.org/10.1080/14719037.2019.1604794

Berman, E. M., & West, J. P. (1997). Managing ethics to improve performance and build trust. *Public Integrity Annual*, *2*(1), 21–29.

Bernstein, E. S. (2012). The Transparency Paradox: A Role for Privacy in Organizational Learning and Operational Control. *Administrative Science Quarterly*, *57*(2), 181–216. https://doi.org/10.1177/0001839212453028

Bevan, G., & Hood, C. (2006). WHAT'S MEASURED IS WHAT MATTERS: TARGETS AND GAMING IN THE ENGLISH PUBLIC HEALTH CARE SYSTEM. *Public Administration*, *84*(3), 517–538. https://doi.org/10.1111/j.1467-9299.2006.00600.x

Bevan, G., & Wilson, D. (2013). Does 'naming and shaming' work for schools and hospitals? Lessons from natural experiments following devolution in England and Wales. *Public Money & Management*, *33*(4), 245–252. https://doi.org/10.1080/09540962.2013.799801

Bhatti, W. A., Waris, S., Zaheer, A., & others. (2011). THE EFFECT OF COMMITMENT AND MOTIVATION ON HUMAN TALENT AND ITS CONTRIBUTION TO ORGANIZATIONAL PERFORMANCE. *Management & Marketing*, *6*(3).

Bifulco, R. (2006). Institutional Change and Coproduction of Public Services: The Effect of Charter Schools on Parental Involvement. *Journal of Public Administration Research and Theory*, *16*(4), 553–576. https://doi.org/10.1093/jopart/muj001

Birhanu, A. G., Gambardella, A., & Valentini, G. (2016). Bribery and investment: Firm-level evidence from A frica and L atin A merica. *Strategic Management Journal*, *37*(9), 1865–1877.

Blazovich, J. L., Smith, K. T., & Smith, M. (2013). Employee-friendly companies and work-life balance: Is there an impact on financial performance and risk level? *Journal of Organizational Culture, Communications and Conflict, Forthcoming*.

Bohte, J., & Meier, K. J. (2000a). Goal displacement: Assessing the motivation for organizational cheating. *Public Administration Review*, *60*(2), 173–182.

Bohte, J., & Meier, K. J. (2000b). Goal Displacement: Assessing the Motivation for Organizational Cheating. *Public Administration Review*, *60*(2), 173–182. https://doi.org/10.1111/0033-3352.00075

Bovaird, T. (2007). Beyond engagement and participation: User and community coproduction of public services. *Public Administration Review*, *67*(5), 846–860.

Bovaird, T., & Loeffler, E. (2012). From engagement to co-production: The contribution of users and communities to outcomes and public value. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, *23*(4), 1119–1138.

Boyne, G. A., James, O., John, P., & Petrovsky, N. (2009). Democracy and Government Performance: Holding Incumbents Accountable in English Local Governments. *The Journal of Politics*, *71*(4), 1273–1284. https://doi.org/10.1017/S0022381609990089

Boyne, G. A., Meier, K. J., Meier, K. J., O'Toole Jr, L. J., Walker, R. M., & others. (2006). *Public service performance: Perspectives on measurement and management*. Cambridge University Press.

Brandsen, T., Steen, T., & Verschuere, B. (2018). *Co-production and co-creation: Engaging citizens in public services*. Taylor & Francis.

Brignall, S., & Modell, S. (2000). An institutional perspective on performance measurement and management in the 'new public sector.' *Management Accounting Research*, *11*(3), 281–306. https://doi.org/10.1006/mare.2000.0136

Burke, F., & Black, A. (1990). Improving organizational productivity: Add ethics. *Public Productivity & Management Review*, 121–133.

Carpenter, D. P., & Krause, G. A. (2012). Reputation and Public Administration. *Public Administration Review*, *72*(1), 26–32. https://doi.org/10.1111/j.1540-6210.2011.02506.x

Case, J. (2000). *Employee theft: The profit killer*. John Case & Associates.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). A Practical Introduction to Regression Discontinuity Designs: Foundations. *Elements in Quantitative and Computational Methods for the Social Sciences*. https://doi.org/10.1017/9781108684606

Chen, S. (2023). Performance feedback, blame avoidance, and data manipulation in the public sector: Evidence from China's official city air quality ranking. *Public Management Review*, 1–21.

Chen, S., & Jia, K. (2021). How local governments prioritize multiple conflicting goals: Beyond the sole-goal perspective. *Public Administration*.

Cho, W., & Melisa, W. D. (2021). Citizen Coproduction and Social Media Communication: Delivering a Municipal Government's Urban Services through Digital Participation. *Administrative Sciences*, *11*(2), 59. https://doi.org/10.3390/admsci11020059

Choi, J., Rhee, M., & Kim, Y.-C. (2019). Performance feedback and problemistic search: The moderating effects of managerial and board outsiderness. *Journal of Business Research*, *102*, 21–33. https://doi.org/10.1016/j.jbusres.2019.04.039

Christensen, J., Dahlmann, C. M., Mathiasen, A. H., Moynihan, D. P., & Petersen, N. B. G. (2018). How Do Elected Officials Evaluate Performance? Goal Preferences, Governance Preferences, and the Process of Goal Reprioritization. *Journal of Public Administration Research and Theory*, *28*(2), 197–211. https://doi.org/10.1093/jopart/muy001

Chun, Y. H. (2005). Goal Ambiguity and Organizational Performance in U.S. Federal Agencies. *Journal of Public Administration Research and Theory*, *15*(4), 529–557. https://doi.org/10.1093/jopart/mui030

Clark, B. Y., Brudney, J. L., & Jang, S.-G. (2013). Coproduction of Government Services and the New Information Technology: Investigating the Distributional Biases. *Public Administration Review*, *73*(5), 687–701. https://doi.org/10.1111/puar.12092

Cyert, R. M., March, J. G., & others. (1963). *A behavioral theory of the firm* (Vol. 2). Englewood Cliffs, NJ.

Davis, A., & Ruddle, K. (2012). Massaging the misery: Recent approaches to fisheries governance and the betrayal of small-scale fisheries. *Human Organization*, *71*(3), 244–254.

Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, *35*(2), 252–279.

DeMatthews, D. E., & Serafini, A. (2019). Do Good Principals Do Bad Things? Examining Bounds Of Ethical Behavior In The Context Of High-Stakes Accountability. *Leadership and Policy in Schools*, 1–20. https://doi.org/10.1080/15700763.2019.1668023

Desmidt, S., & Meyfroodt, K. (2021). How does public disclosure of performance information affect politicians' attitudes towards effort allocation? Evidence from a survey experiment. *Journal of Public Administration Research and Theory*.

Dias, J. J., & Maynard-Moody, S. (2007). For-profit welfare: Contracts, conflicts, and the performance paradox. *Journal of Public Administration Research and Theory*, *17*(2), 189–211.

Dizon-Ross, R. (2020). How does school accountability affect teachers? Evidence from New York City. *Journal of Human Resources*, *55*(1), 76–118.

Downe, J., Cowell, R., & Morgan, K. (2016). What Determines Ethical Behavior in Public Organizations: Is It Rules or Leadership? *Public Administration Review*, *76*(6), 898–909. https://doi.org/10.1111/puar.12562

Downe, J., Grace, C., Martin, S., & Nutley, S. (2010). Theories Of Public Service Improvement: A comparative analysis of local performance assessment frameworks. *Public Management Review*, *12*(5), 663–678. https://doi.org/10.1080/14719031003633201

Dudau, A., Glennon, R., & Verschuere, B. (2019). Following the yellow brick road? (Dis)enchantment with co-design, co-production and value co-creation in public services. *Public Management Review*, *21*(11), 1577–1594. https://doi.org/10.1080/14719037.2019.1653604

Edelenbos, J., van Meerkerk, I., & Koppenjan, J. (2017). The challenge of innovating politics in community self-organization: The case of Broekpolder. *Public Management Review*, *19*(1), 55–73.

Eger III, R. J. (2006). Casting light on shadow government: A typological approach. *Journal of Public Administration Research and Theory*, *16*(1), 125–137.

Fairholm, M. R. (2004). Different Perspectives on the Practice of Leadership. *Public Administration Review*, *64*(5), 577–590. https://doi.org/10.1111/j.1540-6210.2004.00405.x

Fox, S., Spector, P. E., & Miles, D. (2001). Counterproductive work behavior (CWB) in response to job stressors and organizational justice: Some mediator and moderator tests for autonomy and emotions. *Journal of Vocational Behavior*, *59*(3), 291–309.

Frant, H. (1997). Reconsidering the determinants of public authority use. *Journal of Public Administration Research and Theory*, *7*(4), 571–590.

Gao, J. (2021). Mitigating Pernicious Gaming in Performance Management in China: Dilemmas, Strategies and Challenges. *Public Performance & Management Review*, *44*(2), 321–351. https://doi.org/10.1080/15309576.2020.1775662

Gazley, B., LaFontant, C., & Cheng, Y. (Daniel). (2020). Does Coproduction of Public Services Support Government's Social Equity Goals? The Case of U.S. State Parks. *Public Administration Review*, *80*(3), 349–359. https://doi.org/10.1111/puar.13161

Gilad, S. (2015). POLITICAL PRESSURES, ORGANIZATIONAL IDENTITY, AND ATTENTION TO TASKS: ILLUSTRATIONS FROM PRE-CRISIS FINANCIAL

REGULATION: REGULATORY ALLOCATION OF ATTENTION. *Public Administration*, *93*(3), 593–608. https://doi.org/10.1111/padm.12155

Gillam, S. J., Siriwardena, A. N., & Steel, N. (2012). Pay-for-Performance in the United Kingdom: Impact of the Quality and Outcomes Framework--A Systematic Review. *The Annals of Family Medicine*, *10*(5), 461–468. https://doi.org/10.1370/afm.1377

Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel. *Psychological Science*, *20*(3), 393–398. https://doi.org/10.1111/j.1467-9280.2009.02306.x

Girth, A. M. (2014). A Closer Look at Contract Accountability: Exploring the Determinants of Sanctions for Unsatisfactory Contract Performance. *Journal of Public Administration Research and Theory*, *24*(2), 317–348. https://doi.org/10.1093/jopart/mus033

Goddard, M., Mannion, R., & Smith, P. (2000). Enhancing performance in health care: A theoretical perspective on agency and the role of information. *Health Economics*, *9*(2), 95–107. https://doi.org/10.1002/(SICI)1099-1050(200003)9:2<95::AID-HEC488>3.0.CO;2-A

Goh, E., & Kong, S. (2018). Theft in the hotel workplace: Exploring frontline employees' perceptions towards hotel employee theft. *Tourism and Hospitality Research*, *18*(4), 442–455. https://doi.org/10.1177/1467358416683770

Gonzalez, N., Moll, L. C., Tenery, M. F., Rivera, A., Rendon, P., Gonzales, R., & Amanti, C. (1995). Funds of knowledge for teaching in Latino households. *Urban Education*, *29*(4), 443–470.

Greenberg, J. (1990). Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology*, *75*(5), 561.

Greenberg, J. (1997). *A social influence model of employee theft: Beyond the fraud triangle.*

Greenberg, J. (2002). Who stole the money, and when? Individual and situational determinants of employee theft. *Organizational Behavior and Human Decision Processes*, *89*(1), 985–1003. https://doi.org/10.1016/S0749-5978(02)00039-0

Greve, H. (2003). *Organizational Learning from Performance Feedback*. Cambridge.

Grimmelikhuijsen, S. G. (2012). *Transparency and trust. An experimental study of online disclosure and trust in government* [PhD Thesis]. University Utrecht.

Gupta, A. K., Smith, K. G., & Shalley, C. E. (2006). The Interplay Between Exploration and Exploitation. *Academy of Management Journal*, *49*(4), 693–706. https://doi.org/10.5465/amj.2006.22083026

Han, X., & Wang, W. (2022). Does granting managerial autonomy in exchange for accountability mitigate gaming? *Public Administration Review*, puar.13564. https://doi.org/10.1111/puar.13564

Hassan, S. (2019). We Need More Research on Unethical Leadership Behavior in Public Organizations. *Public Integrity*, *21*(6), 553–556. https://doi.org/10.1080/10999922.2019.1667666

Hassan, S., Wright, B. E., & Yukl, G. (2014). Does Ethical Leadership Matter in Government? Effects on Organizational Commitment, Absenteeism, and Willingness to Report Ethical Problems. *Public Administration Review*, *74*(3), 333–343. https://doi.org/10.1111/puar.12216

Heckman, J. J., Heinrich, C. J., & Smith, J. A. (2002). *The performance of performance standards*. National Bureau of Economic Research Cambridge, Mass., USA.

Heinrich, C. J. (2002). Outcomes-Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness. *Public Administration Review*, *62*(6), 712–725. https://doi.org/10.1111/1540-6210.00253

Heinrich, C. J. (2012). How Credible Is the Evidence, and Does It Matter? An Analysis of the Program Assessment Rating Tool. *Public Administration Review*, *72*(1), 123–134. https://doi.org/10.1111/j.1540-6210.2011.02490.x

Heinrich, C. J., & Marschke, G. (2010). Incentives and their dynamics in public sector performance management systems. *Journal of Policy Analysis and Management*, *29*(1), 183–208. https://doi.org/10.1002/pam.20484

Holbein, J. (2016). Left Behind? Citizen Responsiveness to Government Performance Information. *American Political Science Review*, *110*(2), 353–368. https://doi.org/10.1017/S0003055416000071

Holm, J. M. (2018). Successful Problem Solvers? Managerial Performance Information Use to Improve Low Organizational Performance. *Journal of Public Administration Research and Theory*, *28*(3), 303–320. https://doi.org/10.1093/jopart/muy017

Hong, S. (2019a). A Behavioral Model of Public Organizations: Bounded Rationality, Performance Feedback, and Negativity Bias. *JPART*.

Hong, S. (2019b). A Behavioral Model of Public Organizations: Bounded Rationality, Performance Feedback, and Negativity Bias. *Journal of Public Administration Research and Theory*, *29*(1), 1–17. https://doi.org/10.1093/jopart/muy048

Hong, S., Kim, S. H., & Son, J. (2020a). Bounded rationality, blame avoidance, and political accountability: How performance information influences management quality. *Public Management Review*, *22*(8), 1240–1263. https://doi.org/10.1080/14719037.2019.1630138

Hong, S., Kim, S. H., & Son, J. (2020b). Bounded rationality, blame avoidance, and political accountability: How performance information influences management quality. *Public Management Review*, *22*(8), 1240–1263. https://doi.org/10.1080/14719037.2019.1630138

Hong, S., & Kim, Y. (2019). Loyalty or Competence: Political Use of Performance Information and Negativity Bias. *Public Administration Review*, *79*(6), 829–840. https://doi.org/10.1111/puar.13108

Hood, C. (2007). What happens when transparency meets blame-avoidance? *Public Management Review*, *9*(2), 191–210. https://doi.org/10.1080/14719030701340275

Hoover-Dempsey, K. V., Bassler, O. C., & Brissie, J. S. (1987). Parent involvement: Contributions of teacher efficacy, school socioeconomic status, and other school characteristics. *American Educational Research Journal*, *24*(3), 417–435.

Huizinga, K., & de Bree, M. (2021). Exploring the Risk of Goal Displacement in Regulatory Enforcement Agencies: A Goal-Ambiguity Approach. *Public Performance & Management Review*, *44*(4), 868–898. https://doi.org/10.1080/15309576.2021.1881801

Iyer, D. N., & Miller, K. D. (2008). Performance feedback, slack, and the timing of acquisitions. *Academy of Management Journal*, *51*(4), 808–822.

Jacobsen, C. B., & Bøgh Andersen, L. (2017). Leading public service organizations: How to obtain high employee self-efficacy and organizational performance. *Public Management Review*, *19*(2), 253–273. https://doi.org/10.1080/14719037.2016.1153705

Jakobsen, M. (2013). Can Government Initiatives Increase Citizen Coproduction? Results of a Randomized Field Experiment. *Journal of Public Administration Research and Theory*, *23*(1), 27–54. https://doi.org/10.1093/jopart/mus036

Jakobsen, M., & Andersen, S. C. (2013). Coproduction and Equity in Public Service Delivery. *Public Administration Review*, *73*(5), 704–713. https://doi.org/10.1111/puar.12094

Jakobsen, M. L., Baekgaard, M., Moynihan, D. P., & van Loon, N. (2018). Making Sense of Performance Regimes: Rebalancing External Accountability and Internal Learning. *Perspectives on Public Management and Governance*, *1*(2), 127–141. https://doi.org/10.1093/ppmgov/gvx001

Ji, S. (2022). Role of Managerial and Political Accountability in Improving Performance of Local Governments. *Public Performance & Management Review*, *45*(3), 499–531. https://doi.org/10.1080/15309576.2022.2062608

Johnson, W. C., Xie, W., & Yi, S. (2014). Corporate fraud and the value of reputations in the product market. *Journal of Corporate Finance*.

Joseph, J., Klingebiel, R., & Wilson, A. J. (2016). Organizational Structure and Performance Feedback: Centralization, Aspirations, and Termination Decisions. *Organization Science*, *27*(5), 1065–1083. https://doi.org/10.1287/orsc.2016.1076

Julnes, P. de L., & Holzer, M. (2001). Promoting the Utilization of Performance Measures in Public Organizations: An Empirical Study of Factors Affecting Adoption and Implementation. *Public Administration Review*, *61*(6), 693–708. https://doi.org/10.1111/0033-3352.00140

Kang, S., & Van Ryzin, G. G. (2019). Coproduction and trust in government: Evidence from survey experiments. *Public Management Review*, *21*(11), 1646–1664. https://doi.org/10.1080/14719037.2019.1619812

Kelman, S., & Friedman, J. N. (2009). Performance Improvement and Performance Dysfunction: An Empirical Examination of Distortionary Impacts of the Emergency Room Wait-Time Target in the English National Health Service. *Journal of Public Administration Research and Theory*, *19*(4), 917–946. https://doi.org/10.1093/jopart/mun028

Khaltar, O., & Moon, M. J. (2020). Effects of Ethics and Performance Management on Organizational Performance in the Public Sector. *Public Integrity*, *22*(4), 372–394. https://doi.org/10.1080/10999922.2019.1615163

Kiewiet, D. R., & McCubbins, M. D. (2014). State and local government finance: The new fiscal ice age. *Annual Review of Political Science*, *17*, 105–122.

Kim, S. (2004). Individual-Level Factors and Organizational Performance in Government Organizations. *Journal of Public Administration Research and Theory*, *15*(2), 245–261. https://doi.org/10.1093/jopart/mui013

Kogan, V., Lavertu, S., & Peskowitz, Z. (2016). Performance Federalism and Local Democracy: Theory and Evidence from School Tax Referenda: PERFORMANCE FEDERALISM. *American Journal of Political Science*, *60*(2), 418–435. https://doi.org/10.1111/ajps.12184

Koning, P., & Heinrich, C. J. (2013). Cream-skimming, parking and other intended and unintended effects of high-powered, performance-based contracts. *Journal of Policy Analysis and Management*, *32*(3), 461–483.

Kroll, A., & Vogel, D. (2021). Why public employees manipulate performance data: Prosocial impact, job stress, and red tape. *International Public Management Journal*, *24*(2), 164–182. https://doi.org/10.1080/10967494.2021.1874579

Kuenzi, M., Brown, M. E., Mayer, D. M., & Priesemuth, M. (2019). Supervisor-Subordinate (Dis)agreement on Ethical Leadership: An Investigation of its Antecedents and Relationship to Organizational Deviance. *Business Ethics Quarterly*, *29*(1), 25–53. https://doi.org/10.1017/beq.2018.14

Kuusela, P., Keil, T., & Maula, M. (2017). Driven by aspirations, but in what direction? Performance shortfalls, slack resources, and resource-consuming vs. resource-freeing

organizational change: Driven by Aspirations, but in What Direction. *Strategic Management Journal*, *38*(5), 1101–1120. https://doi.org/10.1002/smj.2544

Lachmund, J. (1998). Between scrutiny and treatment: Physical diagnosis and the restructuring of 19th century medical practice. *Sociology of Health & Illness*, *20*(6), 779–801.

Ladd, H. F., & Lauen, D. L. (2010). Status versus growth: The distributional effects of school accountability policies. *Journal of Policy Analysis and Management*, *29*(3), 426–450.

Landi, S., & Russo, S. (2022). Co-production 'thinking' and performance implications in the case of separate waste collection. *Public Management Review*, *24*(2), 301–325. https://doi.org/10.1080/14719037.2020.1823726

Lavie, D., Stettner, U., & Tushman, M. L. (2010). Exploration and Exploitation Within and Across Organizations. *Academy of Management Annals*, *4*(1), 109–155. https://doi.org/10.5465/19416521003691287

Lee, G., & Jimenez, B. S. (2011). Does Performance Management Affect Job Turnover Intention in the Federal Government? *The American Review of Public Administration*, *41*(2), 168–184. https://doi.org/10.1177/0275074010368991

Lehman, D. W., & Ramanujam, R. (2009). Selectivity in organizational rule violations. *Academy of Management Review*, *34*(4), 643–657.

Leone, R., Walker, C., Curry, L. C., & Agee, E. (2012). Application of a marketing concept to patient-centered care: Co-producing health with heart failure patients. *Online Journal of Issues in Nursing*, *17*(2).

Levine, C. H. (2018). Citizenship and service delivery: The promise of coproduction. In *Democracy, bureaucracy, and the study of administration* (pp. 280–300). Routledge.

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, *14*(S2), 95–112. https://doi.org/10.1002/smj.4250141009

Levinthal, D., & March, J. G. (1981). A model of adaptive organizational search. *Journal of Economic Behavior & Organization*, *2*(4), 307–333. https://doi.org/10.1016/0167-2681(81)90012-3

Lewandowski, M. (2019). Public managers' perception of performance information: The evidence from Polish local governments. *Public Management Review*, *21*(7), 988–1010. https://doi.org/10.1080/14719037.2018.1538425

Lindenmeier, J., Seemann, A.-K., Potluka, O., & von Schnurbein, G. (2021). Co-production as a driver of client satisfaction with public service organizations: An analysis of German day-care centres. *Public Management Review*, *23*(2), 210–232. https://doi.org/10.1080/14719037.2019.1674366

Lu, Y., Yang, K., & Thomas, M. B. (2021). Designing Performance Systems in Anticipation of Unintended Consequences: Experiences and Lessons From the Compstat-Based Performance Regime in NYPD. *Administration & Society*, *53*(6), 907–936. https://doi.org/10.1177/0095399720976532

Lucas, G. J. M., Knoben, J., & Meeus, M. T. H. (2018). Contradictory yet Coherent? Inconsistency in Performance Feedback and R&D Investment Change. *Journal of Management*, *44*(2), 658–681. https://doi.org/10.1177/0149206315584821

Lynn, L. E. (1996). *Public management as art, science, and profession*. Chatham House Publishers Chatham, NJ.

Lyons, S. T., Duxbury, L. E., & Higgins, C. A. (2006). A Comparison of the Values and Commitment of Private Sector, Public Sector, and Parapublic Sector Employees. *Public Administration Review*, *66*(4), 605–618. https://doi.org/10.1111/j.1540-6210.2006.00620.x

Magnusson, P. R. (2003). Benefits of involving users in service innovation. *European Journal of Innovation Management*, *6*(4), 228–238. https://doi.org/10.1108/14601060310500940

Maher, C. S., Hoang, T., & Hindery, A. (2020). Fiscal responses to COVID-19: Evidence from local governments and nonprofits. *Public Administration Review*, *80*(4), 644–650.

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87. https://doi.org/10.1287/orsc.2.1.71

March, J. G., & Simon, H. A. (1958). *Cognitive Limits on Rationality.*

Marcus, B., & Schuler, H. (2004). Antecedents of counterproductive behavior at work: A general perspective. *Journal of Applied Psychology*, *89*(4), 647.

Marschall, M. (2006). Parent Involvement and Educational Outcomes for Latino Students. *Review of Policy Research*, *23*(5), 1053–1076. https://doi.org/10.1111/j.1541-1338.2006.00249.x

McMillan, R. (1999). *Competition, parental involvement, and public school performance*. Stanford University.

Meier, K. J., Favero, N., & Zhu, L. (2015). Performance Gaps and Managerial Decisions: A Bayesian Decision Theory of Managerial Action. *Journal of Public Administration Research and Theory*, *25*(4), 1221–1246. https://doi.org/10.1093/jopart/muu054

Meier, K. J., & O'Toole, L. J. (2003). Public Management and Educational Performance: The Impact of Managerial Networking. *Public Administration Review*, *63*(6), 689–699. https://doi.org/10.1111/1540-6210.00332

Meier, K. J., & O'Toole, L. J. (2013). I Think (I Am Doing Well), Therefore I Am: Assessing the Validity of Administrators' Self-Assessments of Performance. *International*

Public Management Journal, *16*(1), 1–27. https://doi.org/10.1080/10967494.2013.796253

Meijer, A. (2013). Co-production in an information age. In *New Public Governance, the Third Sector, and Co-Production* (pp. 210–226). Routledge.

Meijer, A. J., Curtin, D., & Hillebrandt, M. (2012). Open government: Connecting vision and voice. *International Review of Administrative Sciences*, *78*(1), 10–29. https://doi.org/10.1177/0020852311429533

Menzel, D. C. (2015). The ethics factor in local government: An empirical analysis. In *Ethics and Public Administation* (pp. 191–204). Routledge.

Menzel, D. C. (2016). State of the art of empirical research on ethics and integrity in governance. *Ethics in Public Management*, 24–54.

Mikkelsen, M. F., Petersen, N. B. G., & Bjørnholt, B. (2022). Broadcasting good news and learning from bad news: Experimental evidence on public managers' performance information use. *Public Administration*, *100*(3), 759–777. https://doi.org/10.1111/padm.12762

Min, B. H., Oh, Y., & Brower, R. S. (2021). The Effects of Diverse Feedback Dynamics on Performance Improvement: A Typology of Performance Feedback Signals. *Administration & Society*, *53*(1), 123–157. https://doi.org/10.1177/0095399720933822

Minchin, M., Roland, M., Richardson, J., Rowark, S., & Guthrie, B. (2018). Quality of Care in the United Kingdom after Removal of Financial Incentives. *New England Journal of Medicine*, *379*(10), 948–957. https://doi.org/10.1056/NEJMsa1801495

Miner, A. S., Kim, J.-K., Holzinger, I. W., & Haunschild, P. (1996). Fruits of failure: Organizational failure and population level learning. *Academy of Management Proceedings*, *1996*(1), 239–243.

Moynihan, D. P. (2009). Through A Glass, Darkly: Understanding the Effects of Performance Regimes. *Public Performance & Management Review*, *32*(4), 592–603. https://doi.org/10.2753/PMR1530-9576320409

Moynihan, D. P., Fernandez, S., Kim, S., LeRoux, K. M., Piotrowski, S. J., Wright, B. E., & Yang, K. (2011). Performance Regimes Amidst Governance Complexity. *Journal of Public Administration Research and Theory*, *21*(Supplement 1), i141–i155. https://doi.org/10.1093/jopart/muq059

Muller, J. Z. (2018). The tyranny of metrics. In *The Tyranny of Metrics*. Princeton University Press.

Nabatchi, T., Sancino, A., & Sicilia, M. (2017). Varieties of Participation in Public Services: The Who, When, and What of Coproduction. *Public Administration Review*, *77*(5), 766–776. https://doi.org/10.1111/puar.12765

Nederhand, J., & Van Meerkerk, I. (2018). Activating citizens in Dutch care reforms: Framing new co-production roles and competences for citizens and professionals. *Policy & Politics*, *46*(4), 533–550. https://doi.org/10.1332/030557317X15035697297906

Nicholson-Crotty et al. (2017). *Performance and Management in the Public Sector: Testing a Model of Relative Risk Aversion:* 12.

Nielsen, P. A. (2014a). LEARNING FROM PERFORMANCE FEEDBACK: PERFORMANCE INFORMATION, ASPIRATION LEVELS, AND MANAGERIAL PRIORITIES: Learning from performance feedback. *Public Administration*, *92*(1), 142–160. https://doi.org/10.1111/padm.12050

Nielsen, P. A. (2014b). LEARNING FROM PERFORMANCE FEEDBACK: PERFORMANCE INFORMATION, ASPIRATION LEVELS, AND

MANAGERIAL PRIORITIES: Learning from performance feedback. *Public Administration*, *92*(1), 142–160. https://doi.org/10.1111/padm.12050

Nielsen, P. A., & Baekgaard, M. (2015). Performance Information, Blame Avoidance, and Politicians' Attitudes to Spending and Reform: Evidence from an Experiment. *Journal of Public Administration Research and Theory*, *25*(2), 545–569. https://doi.org/10.1093/jopart/mut051

O'Brien, D. T., Offenhuber, D., Baldwin-Philippi, J., Sands, M., & Gordon, E. (2016). Uncharted Territoriality in Coproduction: The Motivations for 311 Reporting. *Journal of Public Administration Research and Theory*, muw046. https://doi.org/10.1093/jopart/muw046

O'Brien, J. P., & David, P. (2014). Reciprocity and R&D search: Applying the behavioral theory of the firm to a communitarian context. *Strategic Management Journal*, *35*(4), 550–565. https://doi.org/10.1002/smj.2105

Osborne, S. P., Strokosch, K., & Radnor, Z. (2018). Co-Production and the Co-Creation of Value in Public Services. In T. Brandsen, T. Steen, & B. Verschuere, *Co-Production and Co-Creation* (1st ed., pp. 18–26). Routledge. https://doi.org/10.4324/9781315204956-3

Ostrom, E. (1972). Metropolitan reform: Propositions derived from two traditions. *Social Science Quarterly*, 474–493.

Owusu, G. M. Y., Koomson, T. A. A., Alipoe, S. A., & Kani, Y. A. (2022). Examining the predictors of fraud in state-owned enterprises: An application of the fraud triangle theory. *Journal of Money Laundering Control*, *25*(2), 427–444. https://doi.org/10.1108/JMLC-05-2021-0053

Parks, R. B., Baker, P. C., Kiser, L., Oakerson, R., Ostrom, E., Ostrom, V., Percy, S. L., Vandivort, M. B., Whitaker, G. P., & Wilson, R. (1981). CONSUMERS AS

COPRODUCERS OF PUBLIC SERVICES: SOME ECONOMIC AND INSTITUTIONAL CONSIDERATIONS. *Policy Studies Journal*, *9*(7), 1001–1011. https://doi.org/10.1111/j.1541-0072.1981.tb01208.x

Parrado, S., Van Ryzin, G. G., Bovaird, T., & Löffler, E. (2013). Correlates of Co-production: Evidence From a Five-Nation Survey of Citizens. *International Public Management Journal*, *16*(1), 85–112. https://doi.org/10.1080/10967494.2013.796260

Percy, S. L. (1978). Conceptualizing and measuring citizen co-production of community safety. *Policy Studies Journal*, *7*, 486.

Pestoff, V. (2006). Citizens and co-production of welfare services: Childcare in eight European countries. *Public Management Review*, *8*(4), 503–519. https://doi.org/10.1080/14719030601022882

Peters, S., & Maniam, B. (2016). Corporate fraud and employee theft: Impacts and costs on business. *Journal of Business and Behavioral Sciences*, *28*(2), 104.

Portz, J. (2021). "Next-Generation" Accountability? Evidence From Three School Districts. *Urban Education*, *56*(8), 1297–1327. https://doi.org/10.1177/0042085917741727

Posen, H. E., Keil, T., Kim, S., & Meissner, F. D. (2018). Renewing Research on Problemistic Search—A Review and Research Agenda. *Academy of Management Annals*, *12*(1), 208–251. https://doi.org/10.5465/annals.2016.0018

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal*, *83*(4), 427–452.

Radin, B. A. (2006). *Challenging the performance movement: Accountability, complexity, and democratic values*. Georgetown University Press.

Rickman, N., & Witt, R. (2007). The Determinants of Employee Crime in the UK. *Economica*, *74*(293), 161–175. https://doi.org/10.1111/j.1468-0335.2006.00530.x

Rustiarini, N. W., T., S., Nurkholis, N., & Andayani, W. (2019). Why people commit public procurement fraud? The fraud diamond view. *Journal of Public Procurement*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/JOPP-02-2019-0012

Rutherford, A., & Meier, K. J. (2015). Managerial goals in a performance-driven system: Theory and empirical tests in higher education. *Public Administration*, *93*(1), 17–33.

Salge, T. O. (2011a). A Behavioral Model of Innovative Search: Evidence from Public Hospital Services. *Journal of Public Administration Research and Theory*, *21*(1), 181–210. https://doi.org/10.1093/jopart/muq017

Salge, T. O. (2011b). A Behavioral Model of Innovative Search: Evidence from Public Hospital Services. *Journal of Public Administration Research and Theory*, *21*(1), 181–210. https://doi.org/10.1093/jopart/muq017

Schafft, K. A., & Brown, D. L. (2000). Social capital and grassroots development: The case of Roma self-governance in Hungary. *Social Problems*, *47*(2), 201–219.

Schmidthuber, L., Ingrams, A., & Hilgers, D. (2021). Government Openness and Public Trust: The Mediating Role of Democratic Capacity. *Public Administration Review*, *81*(1), 91–109. https://doi.org/10.1111/puar.13298

Selin, J. L., & Lewis, D. E. (2018). *Sourcebook of United States executive agencies*.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129.

Simon, H. A. (1966). Theories of decision-making in economics and behavioural science. In *Surveys of economic theory* (pp. 1–28). Springer.

Soehner, D., & Ryan, T. (n.d.). *The Interdependence of Principal School Leadership and Student Achievement*. *5*(3).

Sohn, H., & Bae, K. B. (2018). Do Public Employees "Game" Performance Budgeting Systems? Evidence From the Program Assessment Rating Tool in Korea. *The*

American Review of Public Administration, *48*(5), 458–475. https://doi.org/10.1177/0275074016689322

Svara, J. H. (2008). Beyond Dichotomy: Dwight Waldo and the Intertwined Politics–Administration Relationship. *Public Administration Review*, *68*(1), 46–52. https://doi.org/10.1111/j.1540-6210.2007.00834.x

Taylor, J. (2009). STRENGTHENING THE LINK BETWEEN PERFORMANCE MEASUREMENT AND DECISION MAKING. *Public Administration*, *87*(4), 853–871. https://doi.org/10.1111/j.1467-9299.2009.01788.x

Taylor, J. (2021). Public Officials' Gaming of Performance Measures and Targets: The Nexus between Motivation and Opportunity. *Public Performance & Management Review*, *44*(2), 272–293. https://doi.org/10.1080/15309576.2020.1744454

Thaler, J., & Helmig, B. (2016). Do Codes of Conduct and Ethical Leadership Influence Public Employees' Attitudes and Behaviours? An experimental analysis. *Public Management Review*, *18*(9), 1365–1399. https://doi.org/10.1080/14719037.2015.1103890

Thomsen, M. K., Baekgaard, M., & Jensen, U. T. (2020). The Psychological Costs of Citizen Coproduction. *Journal of Public Administration Research and Theory*, *30*(4), 656–673. https://doi.org/10.1093/jopart/muaa001

Thomsen, M. K., & Jensen, U. T. (2020). Service Professionals' Response to Volunteer Involvement in Service Production. *Journal of Public Administration Research and Theory*, *30*(2), 220–239. https://doi.org/10.1093/jopart/muz028

van den Bekerom, P., van der Voet, J., & Christensen, J. (2021). Are Citizens More Negative About Failing Service Delivery by Public Than Private Organizations? Evidence From a Large-Scale Survey Experiment. *Journal of Public Administration Research and Theory*, *31*(1), 128–149. https://doi.org/10.1093/jopart/muaa027

Van der Voet, J. (2022). Policy preferences in response to negative economic prospects of COVID-19: A survey-experiment among local politicians in four European countries. *Public Administration*, *100*(2), 356–374.

van der Voet, J. (2022). Search in Response to Negative Performance Feedback: Problem-Definition and Solution-Generation. *Journal of Public Administration Research and Theory*, muac008. https://doi.org/10.1093/jopart/muac008

van der Voet, J. (2023). Search in Response to Negative Performance Feedback: Problem-Definition and Solution-Generation. *Journal of Public Administration Research and Theory*, *33*(1), 43–62. https://doi.org/10.1093/jopart/muac008

Van Thiel, S., & Leeuw, F. L. (2002). The performance paradox in the public sector. *Public Performance & Management Review*, *25*(3), 267–281.

Voorberg, W. H., Bekkers, V. J. J. M., & Tummers, L. G. (2015). A Systematic Review of Co-Creation and Co-Production: Embarking on the social innovation journey. *Public Management Review*, *17*(9), 1333–1357. https://doi.org/10.1080/14719037.2014.930505

Walker, R. M., Damanpour, F., & Devece, C. A. (2011). Management Innovation and Organizational Performance: The Mediating Effect of Performance Management. *Journal of Public Administration Research and Theory*, *21*(2), 367–386. https://doi.org/10.1093/jopart/muq043

Wang, W. (2022). How does performance management affect social equity? Evidence from New York City public schools. *Public Administration Review*, puar.13590. https://doi.org/10.1111/puar.13590

Wang, W., & Kim, T. K. (2022). Examining the effects of a performance management reform on employee attitudes and organizational climate. *Public Management Review*, 1–23. https://doi.org/10.1080/14719037.2022.2026095

Wang, W., & Yeung, R. (2019). Testing the Effectiveness of "Managing for Results": Evidence from an Education Policy Innovation in New York City. *Journal of Public Administration Research and Theory*, *29*(1), 84–100. https://doi.org/10.1093/jopart/muy043

Wise, S., Paton, R. A., & Gegenhuber, T. (2012). Value co-creation through collective intelligence in the public sector: A review of US and European initiatives. *Vine*.

Zhu, L., & Rutherford, A. (2019). Managing the Gaps: How Performance Gaps Shape Managerial Decision Making. *Public Performance & Management Review*, *42*(5), 1029–1061. https://doi.org/10.1080/15309576.2019.1568886

Chapter 2: Performance Feedback in Multiple Goals Environment: Goal Prioritization and Effort Substitution

Introduction

Setting strategic goals is highly important to achieve the outcomes that organizations aimed for. It is also observed that if existing strategies seem to improve management activities, at least somewhat, then organizations might not shift their strategies (Nielsen, 2014a; Rutherford & Meier, 2015). Previous literature attributed this behavior to the cognitive limitation of decision-makers, which is bounded rationality pursuing satisfaction rather than maximization (March & Simon, 1958; Simon, 1966). In other words, organizations often do not find alternatives to replace existing strategies when their management results are within the satisfaction range.

Past performance signals provide significant feedback from the environment and stimulate various responses from organizations, affecting their future performances (Cyert et al., 1963). Negative performance feedback caused by failure to meet performance targets is a particularly significant signal, indicating that existing management activities no longer produce satisfactory performance outcomes. It also therefore threatens the job security of its managers and possibly the survival of the organizations (Greve, 2003). This negative feedback pressures organizations to change their strategies. In other words, organizations recognize negative performance feedback as a serious problem that needs to be immediately resolved (Simon, 1966). Strategic changes that occur in response to performance signals are mostly designed to help improve subsequent performances of the organization (Hong, 2019b; Nicholson-Crotty et al., 2017).

However, the reasoning that strategic changes lead to improved subsequent performances has some limitations. One potential limitation of this reasoning is the assumption that the organization has a singular goal (Chen & Jia, 2021; Meier et al., 2015). Furthermore, with this assumption, many prior studies have examined the relationship between performance shortfall and organizational change as default through problemistic search (Posen et al., 2018); a problemistic search continues until a satisfactory level of performance is reached (Cyert et al., 1963; Simon, 1956). In a single-goal context, organizations focus on the failed goal and improving their performance via problemistic searching and organizational change. However, as previous literature noted, one of the key traits of public organizations is goal multiplicity (Boyne et al., 2006; Holm, 2018). Public organizations are often required to seek multiple goals. Thus, in contrast to previous perspectives on performance feedback, this study investigates organizational strategy changes in response to performance shortfalls in multiple-goal environments.

Goal prioritization in response to performance signals is primarily associated with Performance improvement (Holm, 2018; Nielsen, 2014a). Goal prioritization is the transition of organizational attention to highlight various goals which have not been previously paid attention to (Nielsen, 2014a; Rutherford & Meier, 2015). It means that low performing organizations may prioritize the set of goals, which are below the performance target, and input their efforts towards fixing the problem (Cyert et al., 1963; Walker et al., 2011). Goal prioritization therefore enables low-performing organizations to identify the high-stakes goals that they should immediately achieve and assign resources on those goals.

This study, however, focuses on the unintended effect of goal prioritization. Although prioritization of low-performed goals may lead to performance improvement

(Holm, 2018), it could also cause effort substitution that public organizations deliberately pay less attention to low-stakes goals in performance metrics (Benaine & Kroll, 2020; Han & Wang, 2022). With bounded rationality and resource limitations, public organizations may pay less attention to goals which are essential to nurture public values but are less helpful to improve measured performance (Kelman & Friedman, 2009). While effort substitution could be harmful for public organizations to achieve diverse public values, a very limited number of studies previously have focused on this issue. Instead, previous literature has investigated how performance signals affect goal prioritization and the influence of goal ambiguity and accountability mechanism on effort substitution (Holm, 2018; Lu et al., 2021; Nielsen, 2014a; Taylor, 2021). Overall, although goal prioritization may be related to cheating behaviors in public sector, previous literature paid less attention to risk of occurring effort substitution in low performing organizations.

Therefore, in order to remedy this gap in our understanding, this study examined whether there is effort substitution towards high-stake goals in low-performing organizations triggered by negative performance feedback. This study presents empirical evidence that low performing organizations prioritize overall performance outcomes but paid less attention to low-stakes goals. This paper mainly utilizes the School Quality Report of elementary schools of New York City. Education has been an emerging sector to test the theories of performance management because of relatively clear goals and quantifiable performance information. (Holm, 2018; Nielsen, 2014a; Wang, 2022; Wang & Kim, 2022). School Quality Reports are issued annually by the New York City Department of Education (NYCDOE), which allows annual or rapid strategic evaluation. They include four performance grades (Not meeting, Approaching, Meeting, and Exceeding) and continuous

scores (1.00-4.99) based on which grades are assigned. During the study period from 2015 to 2019, performance ratings were published on the NYCDOE website and the media, but there were no formal sanctions on low performers. This gives a unique opportunity for researchers to employ a quasi-experimental regression discontinuity design (RDD) to estimate the impact of performance feedback as each school has both scores and grades.

Furthermore, this paper makes several contributions on theory and practice. It offers novel perspectives on the relationship between performance management and effort substitution. Previous literature on performance management paid limited attention to how different level of performance outcomes affect organizations in their engagement in effort substitution. As an additional contribution, this study deepens our understanding of naming-and-shaming accountability system. Results of this paper indicate that naming-and-shaming pressures low performing organizations to improve performance outcomes, but it may not remedy effort substitution. It gives us a lesson that external accountability may not be a solution for effort substitution, and it is necessary to have alternative attempts including internal learning and managerial autonomy.

Accountability Pressure as Motivation of Performance Improvement

Influenced by New Public Management (NPM), many governments designed their systems to hold the agencies accountable. Those systems often include high-stakes accountability mechanisms based on performance outcomes (Zhu & Rutherford, 2019). The possibility of receiving sanctions is an important aspect of accountability pressure (Andersson & Liff, 2012). Sanction, in this context, indicates negative incentives on organizations where the

performance outcomes were not satisfactory (Girth, 2014). Sanctions from the regulators triggered by low-performance signals of an organization, increase the pressure for the said organization to improve performance outcomes (Greve, 2003; Meier et al., 2015; Nielsen, 2014a; Salge, 2011a). That is, low performance signals of an organization stimulate more fear of the potential or ongoing sanctions from the regulators, and it pressures organizations to improve performance outcomes. There are a wide range of accountability systems deployed to hold public officials accountable. Rewards-and-sanctions system is the prevalent system, which imposes accountability pressure with visible and direct sanctions on organizations with unsatisfactory performance (Bevan & Wilson, 2013). Prior studies tested the impact of low performance signals on performance improvement in the rewards-and-sanction context (Holm, 2018; Hong, 2019b). In this system, both reward and sanction are extended so that high-performing agencies receive rewards such as financial incentives, while low-performing agencies face sanctions including dismissal and budget cuts. Many countries have adopted rewards-and-sanctions accountability mechanisms such as Program Assessment Rating Tools (PART) in United States (Heinrich, 2012), the Korean Performance Assessment Rating Tool (Min et al., 2021; Sohn & Bae, 2018), and hospital performance management in United Kingdom (Kelman & Friedman, 2009).

Additionally, it is important to understand that accountability pressure in low performing organizations is not generated only by rewards-and-sanctions accountability system. Even without visible and direct sanctions on organizations, managers could be pressured by low performance signals. For instance, performance outcomes could become high stakes when politicians governing public organizations are seeking electoral support. If political elections or public voting directly impacts the management, this type of

accountability leans on their reputation damages rather than financial and personnel sanctions. This is commonly called naming-and-shaming (Bevan & Wilson, 2013). In this case, since performance outcomes are closely related to re-election (Boyne et al., 2009), high performance may return high electoral support, which enhances the probability of re-election. On the other hand, a low-performance outcome may lead to failure in re-election. Ji (2022) found that local governments governed by politicians do respond with enhanced sensitivity to low-performance signals when the accountability system relies on naming-and-shaming, supporting this further.

Although naming-and-shaming accountability mechanisms could be relatively low stakes for public managers who are not related to elections, naming-and-shaming accountability mechanism could still provide accountability pressures for underperforming organizations to enhance their overall performance outcomes. This is because information transparency and quality information creates an awareness among the public managers of the possible aspiration level, which is the level of desired performance by organizations (Hong, 2019b; Salge, 2011a). The aspiration level works as a reference point for organizations to compare their current performance with their past performance, as well as comparison with the performance of peer organizations. Through comparison, low-performing organizations initiate a learning process to find the problems and search for alternatives to innovate themselves (Greve, 2003; Nicholson-Crotty et al., 2017). On the other hand, damage to reputation may directly and indirectly affect managers' future careers (Moss and Sanchez 2004). In response to the performance shortfalls, co-workers and the public may question their leadership, which affect co-workers to perceive the reputation of managers negatively. This means that performance shortfalls can be a signal

for a manager's poor competency (Carpenter & Krause, 2012). Negative performance feedback also draws the attention of senior officers and councilors on the management (Downe et al., 2010).

Thus, this leads to the question: If accountability pressure drives underperforming organizations to achieve the desired level of performance, how can they achieve it? The next section discusses goal prioritization as a strategy of underperforming organizations and further their effort substitution. Although effort substitution is the unintended effect of goal prioritization, this could affect improvement of performance outcomes by leading managers to be more responsive on the measured indicators in performance evaluation.

Performance Signals in Multiple Goals Environment: Goal Prioritization

Goal prioritization is to redirect organizational attentions or resources to highlight different goals which have not been paid attention to before (Nielsen, 2014a). Previous literature has identified goal prioritization as a managerial strategy in response to poor performance signals and examined performance improvement through goal prioritization (Chen & Jia, 2021; Holm, 2018). Goal prioritization occurs in the environment where organizations face multiple goals and have different priority rules towards each goal. Goal multiplicity is not a novel trait in organizations (Boyne 2002; Holm 2018). Regardless of whether an organization is private or public, it is likely required to seek multiple goals. The difference is that public organizations often have conflicting goals which causes goal ambiguity (Rainey and Jung 2014). Private organizations have hierarchical goals so that there is a

relatively clear ultimate goal which is often the profit, and there are other subgoals to help achieve the ultimate goal (Mazzelli et al. 2019). However, public organizations often have horizontal goals which makes it challenging to judge the relative importance of goals (Gilad, 2015; Moynihan et al., 2011). For example, public organizations may agonize between achieving budget efficiency and broadening provision of public service, where the former is related to a decrease in citizens' tax burden, while the latter accomplishes equity but leads to an increased tax burden. For that reason, public management literature has been interested in studying how multiple goals affect management and vice versa (Chun and Rainey 2005; Davis and Stazyk 2016).

Goal prioritization derives from bounded rationality and resource limitation of an organization (Cyert et al., 1963; Simon, 1966). This fundamentally means that managers set the goals of organizations with constraints including cognitive and resource limitations (Rutherford & Meier, 2015). First, cognitive limitations of top managers lead organizations to sequentially pursue their goals (Greve, 2003; Hong et al., 2020b). Empirical evidence shows that top managers prefer to sequentially attend to a goal since it reduces their cognitive efforts and allows easier decision making (Argote & Greve, 2007). In this sense, cognitive limitations of top managers hinders targeting multiple goals simultaneously. Resource limitation is another reason why organizations prioritize the goals. Abundance in resources enables multiple options and flexibility to pursue solving a challenge. However, organizations may have insufficient resources to accomplish multiple goals simultaneously. In reality, almost all organizations have insufficient resources to achieve their goals. Therefore, one of the crucial roles of a manager is efficient reallocation of limited resources (Cyert and March 1963; Katsuhiko 2007).

In a multiple goal setting, several conditions could affect priority rules in the process of goal prioritization (Argote & Greve, 2007; Cyert et al., 1963). First, performance gap could be a compelling rule to prioritize organization goals in public management (Chen & Jia, 2021; Nielsen, 2014a; Rutherford & Meier, 2015). Specifically, underperforming organizations tend to prioritize sets of goals which are below their historical performance trend or achievement level of other organizations (Hong, 2019b). Second, requests or pressures from external environment may also influence priority rule in the goal prioritizing process. (Holm, 2018). As public organizations are embedded in the political environment, accountability pressure from citizens and regulators strongly incentivizes public managers to prioritize goals following preference of external actors (Chen & Jia, 2021; Meier et al., 2015). Furthermore, resource constraints may influence the priority rules. In this sense, underperforming organizations are likely to intensively invest resources towards their priority goal(s), since they are more likely to suffer from resource constraints than high performing organizations (Kuusela et al., 2017). This is because, in the public sector, regulators may put sanctions on low performing organizations including cutting the budget of poor-performing organizations, and competent employees may leave these organizations (Lee & Jimenez, 2011; Min et al., 2021). Overall, it could be deduced that negative performance signals drive organizations to prioritize high-stakes goals among their multiple goals (Nielsen, 2014a).

Furthermore, prioritization on low-performed goals positively affect performance improvement since it means that organizations invest resources for analyzing the reasons for poor performance and orientate towards fixing the problem (Cyert et al., 1963; Meier & O'Toole, 2003; Walker et al., 2011). Investing resources on low-performed goals may

turn into performance growth on these goals because of high marginal return (Holm, 2018). When organizations invest the same amount of resources, low-performed goals have relatively more room to improve on outcomes than high-performed goals. This means that when a goal was initially low performed, it requires less effort for an organization to turn into a high-level performance growth. This expectation is examined in the first hypothesis.

Hypothesis 1: In response to poor performance signals, organizations improve their overall performance outcomes.

Unintended Effect of Goal Prioritization: Effort Substitution

Although (re)prioritizing goals and efforts may lead to performance improvement (Holm, 2018), it may also cause public organizations to unintentionally crowd out other goals. These other goals are often essential to nurture public values but are less helpful to improve measured performance. In other words, public organizations may narrow down their scope to focus on a goal that is not only easy to measure, but also has high stakes for improving measured performance; this is called effort substitution (Benaine & Kroll, 2020; Han & Wang, 2022). This difference in goal importance indicates that managers may "treat them as individual items that each enter into a manager's decision calculus" (Meier et al. 2015: 1236). Organizations therefore would pay less attention to a goal that is difficult to measure or is missing from a performance evaluation (Kelman & Friedman, 2009). For example, in public education, schools may put more efforts into achieving high performance in high-stakes subjects, such as Math and English Language Arts (Dee et al., 2013). As another

example, elected officials may prioritize desirable outcomes in order to win a future election, but some low-stake goals are less focused on even if the performance outcome of them is below the aspiration level (Christensen et al., 2018).

Effort substitution is problematic in public management as it narrowly focuses on efficiency and may harm equity and democracy (Radin, 2006). For instance, Koning & Heinrich (2013) observed that in performance-based contract, social benefit administration may prioritize serving unemployed insured workers, who have better chances of finding a job, than disability insured workers. However, effort substitution is different from other strategic behaviors such as distortion and cheating, because this effort may not be purposeful. Rather, it is close to an unintended effect caused by goal prioritization which is legitimate in public management (Benaine & Kroll, 2020). Considering bounded rationality of managers and resource constraints of an organization, effort substitution could be an excuse for behaviors of low performers who need to prioritize high-stakes indictors. That is, effort substitution could be an unintended result from goal prioritization under the environments of accountability pressure, predetermined performance indicators, bounded rationality and resource limitation. Although it is an issue related to social equity, it is also an unintended outcome in a way to hold a performance-based contract.

Thus, as effort substitution is closely related to goal prioritization, past performance signals could be the antecedent of effort substitution. However, this relationship has been relatively overlooked in public management literature. Specifically, the public management literature has focused more on how performance signals affect goal prioritization (e.g. Holm, 2018; Nielsen, 2014). Furthermore, public management literature mostly relies on goal ambiguity and accountability mechanism to explain effort substitution

(Lu et al., 2021; Taylor, 2021). However, the literature on effort substitution has limitations in that it does not explain the diverse patterns of public organizations in response to past performance signals. This could limit the opportunity to understand public organizational goal displacement in depth. Exploring how past performance signals affect effort substitution also enables us to find specific groups that need to be effectively engaged with for mitigating effort substitution. Summarizing the above scenario, this study proposes the following hypothesis:

Hypothesis 2: *In response to performance shortfalls, organizations are less concerned with the low-stakes goal in performance metrics.*

Institutional Background

Similar to the first chapter, evidence for this paper comes from the ratings and scores of the New York City School Quality Report from 2014-15 AY to 2017-18 AY. School Quality Reports provide a summary of key information on a school's practices and performance. This is the primary dataset for this study. The School Quality Report is issued annually by the New York City Department of Education (NYCDOE) and includes performance grades as well as continuous scores based on which grades are assigned.

In School Quality Reports, academic achievement, which is the primary dataset in this study, is measured using outcomes based on state exam and growth results, including Math and English Language Arts (ELA). These exams have different weights on their overall scores. In performance metrics in elementary schools, indicators of Social Studies

and Science are not included in evaluation. Specifically, overall student achievement equals the weighted average score plus additional points that close the achievement gap. Measurement for student achievement in elementary and middle school includes percentages at Levels 3 and 4 in Math and ELA, average scores in Math and ELA test results, and core course pass rates. Students at Levels 3 and 4 indicate that they are proficient and are excelling for their grade. Level 3 and 4 are the basis for measuring the average proficiency rates. Also, when high-need students achieve progress in their tests, additional points are given to the school which can help to close the achievement gap. High-need students in this setting include English Language Learners, students with disabilities, and students of racial minority groups, such as Black and Hispanic ethnicity. However, the portion of this extra point in student achievement metrics is not substantial; extra points can add a maximum of 0.25 points to the overall score; the maximum overall score is 4.99. High school students' achievement measures take into account different dimensions of performance, such as graduation rates, percentage of students earning 10+ credits, Regents exam results, and so on.

The frame of School Quality Reports has been consistent from 2014-15 AY to 2018-19 AY periods. The last time performance measurements were transformed was in 2013-14 AY. There was a significant difference in performance measurements before and after 2013-14 AY which is when NYCDOE withdrew the reward-and-sanction accountability system in their performance management. From 2006-07 AY to 2012-13 AY, NYCDOE linked rewards-and-sanctions with the letter grades. 'A' grade schools received financial rewards, such as performance-related bonuses and extra funds based on the number of students. On the other hand, 'D' or 'F' schools faced sanctions including

leadership change, school restructuring, or closure. For example, in 2008, five schools that received 'F' grades were closed or merged with other schools. After 2014, NYCDOE no longer imposes rewards-and-sanctions based on performance outcomes in public schools. However, performance ratings continue to be published on the NYCDOE website and mass media.

In addition to the policy for rewards-and-sanctions, NYCDOE's performance management experiment in 2014 also entailed changes in performance grades. Performance grades were changed into 4 ratings: Exceeding, Meeting, Approaching, and Not meeting Targets. Before 2013-14 AY, there were five grades, including A, B, C, D, and F. The changed ratings are distributed by performance scores that range from 1.00 to 4.99. The first digit indicates the grade level and the subsequent digits show proximity of the result to the next level. For instance, suppose that school A scored 1.30 and school B scored 1.90; the grades of these two schools would be marked as 'not meeting target' because the scores of both these schools started with '1'. However, school B was closer to the 'approaching target' rating than school A since the subsequent digits of school B (.90) were higher than those of school A (.30).

In the public education context, additionally, effort substitution is possible to occur because of principals' authorities and activities. With instructional leadership, principals have ability to directly and indirectly affect student achievement with the instructional leadership including "the expertise of the classroom teacher interacting with students and actually teaching students how and what to learn (Soehner & Ryan, 2011: 283)." To be specific, principals could be involved in setting instructional strategies, changing teaching schedules and curriculum, designing after-school programs, and evaluating student

progress (Glasman, 1984; Nettles & Herrington, 2007). Those activities may allow principals to focus on certain indicators in the performance evaluation.

Data and Method

Dependent Variable

This study leverages student achievement scores to examine whether the performance shortfalls positively influence the improvement of overall performance. This paper also utilizes average proficiency rate of Science as a variable for underrepresented goal. Dependent variables are measured one year after receiving performance grades. This study relies on dataset from public elementary schools in New York City. This is because the evaluation system for elementary schools does not include Social Studies and Science whereas the performance metrics for middle schools include scores of Social Studies and Science as indicators of student achievement evaluation. However, for the statistical power, this study adds the analyses with data of middle schools for robustness check. Hence, the average proficiency rate of Science is used as an unmeasured and low-stakes goal in the analyses. Similar to the average proficiency rates in Math and ELA, the average proficiency rate in Science indicates the percentage of students who achieved Levels 3 and 4 in Science. Because proficiency rate for Social Science is not available in School Report Card during the study period, this study only utilizes Science proficiency rate.

Empirical Strategy

The School Quality Report dataset gives a unique opportunity for researchers to estimate the impact of performance feedback. The public schools performance is evaluated as a continuous score, and the results are published as four grades. This setting, where NYCDOE assigns performance grades to schools, allows this study to use Regression Discontinuity (RD) design as the primary methodology; RD enables a comparison between schools where the performance is just above the grade cutoff and those where it is just below the grade cutoff. This design provides a credible causal inference about how performance affects subsequent organizational changes. To utilize an "sharp" RD design, there are three required features: 1- scores - which are called a running variable, 2- cutoffs, and 3- treatment. Specifically, the score is continuously distributed, the cutoff point is clear-cut, and all observations comply with the treatment assignment (Cattaneo et al., 2019). The School Quality Report dataset meets the criteria for a RD design; score is distributed from 1.00 to 4.99, grades are clear-cut, and all observations follow the cutoff rule.

RD relies on the assumption that units near the cutoff are similar to each other, except for the treatment status (Cattaneo et al. 2019). For example, schools with performance scores between 1.99 and 2.00 are unlikely to have significant differences; 0.01 would not likely to be significant difference in performance level. It could be deduced that these schools are less likely to have a meaningful difference in their performance. However, the grades 'approaching' for schools at score 2.00 and 'not meeting' for schools at score 1.99 generate different signals so that stakeholders, including staff, teachers, principals, and parents, may treat them differently. While stakeholders of schools with an 'approaching' grade may perceive it as a signal for the need for an incremental change to achieve better performance in the future, stakeholders of schools with a 'not meeting' grade

may perceive it as a serious signal for the need for a radical change. Therefore, differences of performance grade near the cutoff may be attributed to performance signals.

Furthermore, RD can overcome limitations of the Ordinary Least Squares (OLS) model. A legitimate concern in the OLS model is that high-performing organizations may have fundamental differences from low-performing organizations, but the OLS model does not effectively control these unobservable factors (Hong et al., 2020b). This problem therefore motivates the usage of a RD design, which assumes that the distribution of unobservable factors that cause biases in the OLS model changes smoothly across performance grades. Based on this assumption, the RD design examines whether a significant difference exists among organizations that are just above and just below the cutoff.

Regression Discontinuity Design

This study employs Local Polynomial Regression Discontinuity (*hereafter, local RD*) that is non-parametric when investigating the relationship between outcome variables and a running variable. Whereas the global polynomial RD uses continuity of average to compare just below and just above the cutoff, the local RD utilizes bandwidth[5] to estimate the RD point estimator (Cattoneo et al. 2019). Hence, the local RD model estimates the RD results with observations around the cutoff point so that it enables us to measure the changes occurring near the cutoff point. In this setting, the criterion of generating bandwidth

[5] Bandwidth "determines the size of the neighborhood around the cutoff where the empirical RD analysis" (Cattoneo et al. 2019: 41).

influences the reliability of estimators because arbitrary bandwidth could raise concerns of specification searching and *ad-hoc* decisions; this threatens the validity and reliability of the results. For this reason, this study employs several bandwidths introduced by Cattaneo et al. (2019). Specifically, this study adopts three commonly-used bandwidths: 1- Mean squared error (MSE)[6] optimal, 2- MSE sum, and 3- Coverage error (CER)[7] optimal. Each bandwidth selection minimizes the mean squared error, mean squared error of the sum of the regression coefficient, and the asymptotic coverage error rate, respectively. In addition, this study also adopts a local quadratic polynomial with MSE-optimal to test the robustness of the result. The local RD model employed in this study is as follows:

$$Y_{s(t+1)} = a + f\left(X_{s(t)}\right) + \beta G_{s(t)} + \gamma Covariates_t + \pi_t + \epsilon_{st}$$

Where $Y_{s(t+1)}$ indicates the dependent variables, which indicate overall performance and the low-stakes goal in performance metrics. Specifically, they are the student achievement scores and Science proficiency rates of school, s, in year t+1. $f(X_{st})$ is a linear and quadratic polynomial of the running variable X, which is continuously distributed performance scores, in year t, and $G_{s(t)}$ is the grade determined by the running variable. The vector of covariates consists of the number of enrollments, races (White, Black, and Hispanic), Economic Need Index, English language learner (ELL), and Student with

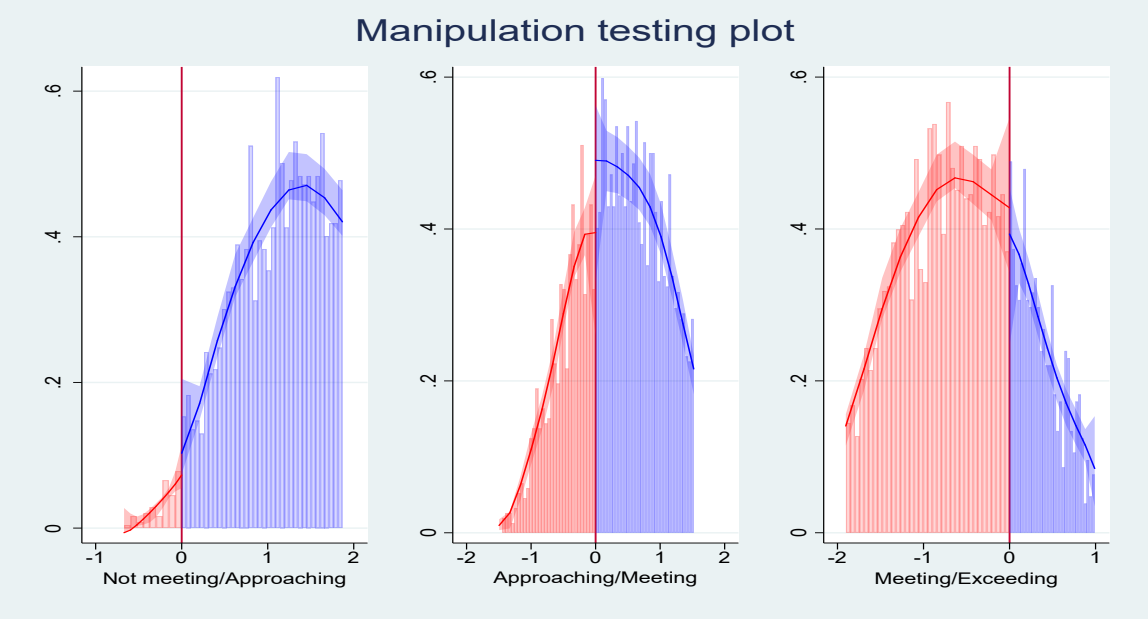
[6] Mean squared error occurs when there is an average squared difference between the actual and estimated values.
[7] Coverage error occurs when the actual sample does not correspond to the target population.

Disability (SWD). Further, this model includes teacher response rates as a covariate. Π_t is the time fixed effects.

Validation tests for Regression Discontinuity Design

The validation of RD design is based on three assumptions. First, units have no capacity to manipulate their scores, which is called no manipulation of the treatment. It means that schools are not able to deliberately alter the value of their performance scores to increase or decrease their overall performance and proficiency rates. Second, the cutoff point should be unique in the running variable, which means there should be no significant changes at other cutoff points. To satisfy this assumption, the trend of outcome variables would not be significantly changed at the other cutoff points. Third, variables except for the dependent variables show no systematic difference between the treatment and control groups. Thus, to show that RD design in this study fulfill these assumptions, density test, placebo test, and balanced check have been done.

Firstly, this paper employs density tests to check whether there are significant changes in density at the cutoff point. If schools have capacity to manipulate the running variables, the density of schools would show significant changes at the cutoff points. Figure 1 graphically indicates that there are no statistically significant jumps or drops at the cutoff point. Instead, the trends of density smoothly continued over the cutoffs. Table 2 confirms the results from this graph. There are no p-values that are smaller than 0.05 and no t-values that are higher than this threshold for statistical significance. The results support that schools did not manipulate the treatment status.



Note: Line and shading area indicate the point estimate and 95% confidence interval respectively

Figure 1 Density Plots at Cutoffs

Table 2 Manipulation Test Results

Cutoff	(1) Not meeting/Approaching	(2) Approaching/Meeting	(3) Meeting/Exceeding
T-value	1.44	0.92	-1.35
P-value	0.15	0.36	0.18
Effective N	513	1390	1415

Note: The null hypothesis is that there is no difference in the density of treated and control observations at the cutoff

To check the second assumption, this paper employs placebo tests. A placebo test runs the main model with alternative cutoffs, which is close to but different from the actual cutoffs. Results with alternative cutoffs should show no significant discontinuities at these cutoffs. This test focuses on the cutoff between 'not meeting target' and 'approaching

target', which is the interests of this study. This paper sets the alternative cutoffs where the scores are 1.92, 1.96, 2.04, and 2.08 for the true cutoff of '2' that is the cutoff point for 'not meeting target' grade and 'approaching target' grade. The placebo test is conducted only by using observations above the true cutoff and control variables for points below the true cutoff (Cattaneo et al., 2019). Table 3 shows the results of the placebo test. The values of dependent variables including student achievement and Science proficiency rate smoothly continued over the alternative cutoffs. These results support that the second assumption is not violated in the RD model.

Table 3 Results with placebo cutoffs

	(1)	(2)
	Subsequent Student Achievement	Science Proficiency Rate
Alternative cutoff : **-0.08**	0.20	7.58
	[-0.91, 0.70]	[-2.46, 26.14]
Effective N	40	35
Alternative cutoff : **-0.04**	0.08	3.33
	[-0.48, 0.20]	[-12.94, 27.2]
Effective N	30	31
Alternative cutoff : **+0.04**	0.33	2.42
	[-0.10, 0.77]	[-5.48, 18.43]
Effective N	156	143
Alternative cutoff : **+0.08**	-0.09	2.90
	[-0.70, 0.17]	[-4.57, 13.72]
Effective N	173	162

Finally, this paper checks the balance with predetermined variables. The balance check is to measure if the control and treatment groups have similar characteristics near the cutoff points (Cattaneo et al., 2019). To verify that there are no systematic differences, this study

utilized predetermined covariates in 2012-13 AY, which are the data before the study period, and before the performance regime change in 2013-14 AY, and ran the RD model. The predetermined covariates were not overlapped with the main model which is based on the data from 2014-15 AY to 2018-19 AY.

In Table 5, although the percentages of Black and Hispanic students show weak significance at 90% confidence level, we could say that the observations near the cutoff have similar characteristics because the conventional threshold for statistical significance is 95% level. The values of predetermined covariates smoothly continued over each cutoff between 'not meeting target' and 'approaching target', between 'approaching target' and 'meeting target', and between 'meeting target' and 'exceeding target'. Overall, the RD design of this study satisfies the assumptions for validation.

Table 5 RD Regression Results on Predetermined Variables (Based on MSE-optimal bandwidth and AY 2012-13)

	(1) Enrollment	(2) Economic Need Index (%)	(3) Black (%)	(4) Hispanic (%)	(5) White (%)	(6) English language learner (%)	(7) Students with disability (%)
RD estimator	36.19	0.03	-2.93*	2.53*	0.68	0.00	0.05
(Not meeting/ Approaching)	[-76.44, 166.03]	[-0.03, 0.13]	[-7.22, 0.24]	[-0.13, 6.73]	[-.53, 2.42]	[-0.02, 0.02]	[-2.76, 2.63]
Effective N	134	107	137	151	164	137	147

RD estimator	1.05	0.00	-0.01	-0.39	-0.28	0.00	-0.13
(Approaching/ Meeting)	[-29.17, 34.06]	[-0.03, 0.02]	[-1.17, 1.45]	[-1.71, 0.58]	[-1.24, 0.44]	[-0.01, 0.01]	[-1.12, 0.92]
Effective N	1187	1138	1166	1213	1256	1088	1256
RD estimator	-8.73	0.01	0.46	-0.85	0.21	0.00	-0.25
(Meeting/ Exceeding)	[-53.7, 27.1]	[-0.02, 0.03]	[-0.93, 1.52]	[-2.07, 0.45]	[-0.59, 1.41]	[-0.01, 0.01]	[-1.52, 1.09]
Effective N	739	875	967	1091	917	739	810

Note: Inference based on robust bias corrected confidence intervals that are presented in brackets. All models include year fixed effects and covariates. Standard errors clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Results

Descriptive Statistics

Table 4 presents the summary statistics of the outcome and control variables. The mean of outcome variables shows stable degrees of student achievement and proficiency rate in Science over the four-year period. The average score of student achievement ranged from 3.26 to 3.58 while the percentage of students who achieved Level 3 and 4 in Science ranged from 84% to 87%.

Regarding control variables, the averages of those variables are stable, similar to the outcome variables. Enrollment decreases from 650 to 588 in the study period, and the Economic Need Index varies from 0.63 to 0.73. The students in New York City Public Schools consist of approximately 30% Black students, 40% Hispanic students, and 15% White students. During the study period, about 14% and 20% of students were regarded as

English Language Learner and Student with Disability. The number of observations ranged from 811 to 817 over the years.

Table 4 Descriptive statistics

		2015-16 AY (N = 814)	2016-17 AY (N = 817)	2017-18 AY (N=811)	2018-19 AY (N = 817)
Dependent variable					
Subsequent Student Achievement	Mean	3.58 (0.72)	3.26 (0.79)	3.34 (0.76)	3.36 (0.77)
Science Proficiency Rate	Mean	86.97 (11.43)	84.17 (13.02)	87.16 (12.11)	84.47 (13.15)
Control variable					
Enrollment	Mean	649.15 (310.84)	607.54 (303.43)	600.46 (298.82)	587.79 (290.76)
Economic need index	Mean	0.66 (0.22)	0.66 (0.22)	0.63 (0.23)	0.73 (0.22)
African American student(%)	Mean	0.31 (0.29)	0.30 (0.29)	0.30 (0.29)	0.30 (0.28)
Hispanic student (%)	Mean	0.40 (0.26)	0.40 (0.26)	0.40 (0.26)	0.41 (0.26)
White student (%)	Mean	0.14 (0.22)	0.15 (0.21)	0.15 (0.21)	0.15 (0.21)
English language learner (%)	Mean	0.14 (0.12)	0.14 (0.11)	0.14 (0.12)	0.15 (0.12)
Students with disability (%)	Mean	0.20 (0.07)	0.20 (0.07)	0.21 (0.07)	0.21 (0.07)

Figure 2 shows the distribution of performance grade in the research period. The distribution seems to have a bell shape for each year. In 2014-15 AY, 1%, 25%, 55%, and 19% of schools received 'not meeting', 'approaching', 'meeting', and 'exceeding target' grades, respectively. In 2015-16 AY, 1%, 21%, 46%, and 32% of schools were regarded

as ‘not meeting’, ‘approaching’, ‘meeting’, and ‘exceeding target’ grades. In 2016-17 AY and 2017-18 AY, 3% and 4% of schools received ‘not meeting target’ grade, 35% and 32% of schools were regarded as ‘approaching target’ grade, 43% and 44% of schools were categorized as ‘meeting target’ grade, and 19% and 22% of schools achieved ‘exceeding target’ grade.

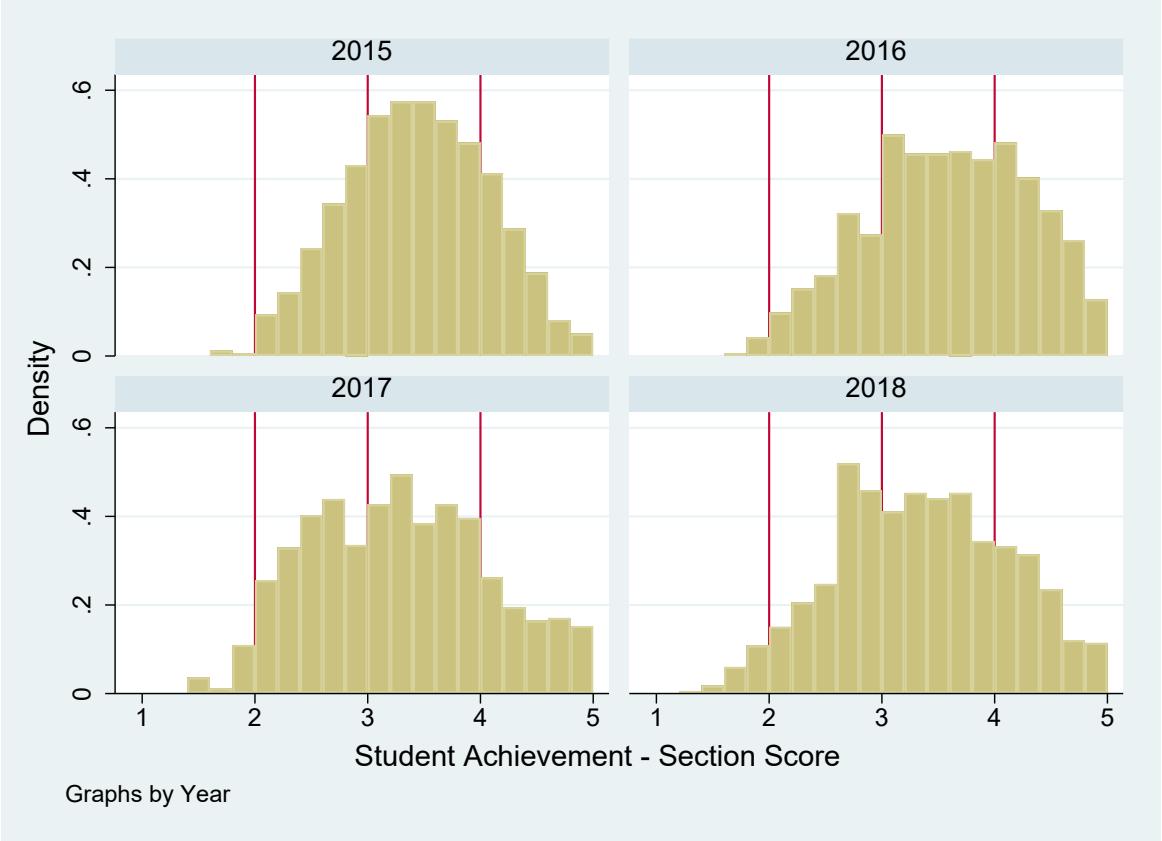
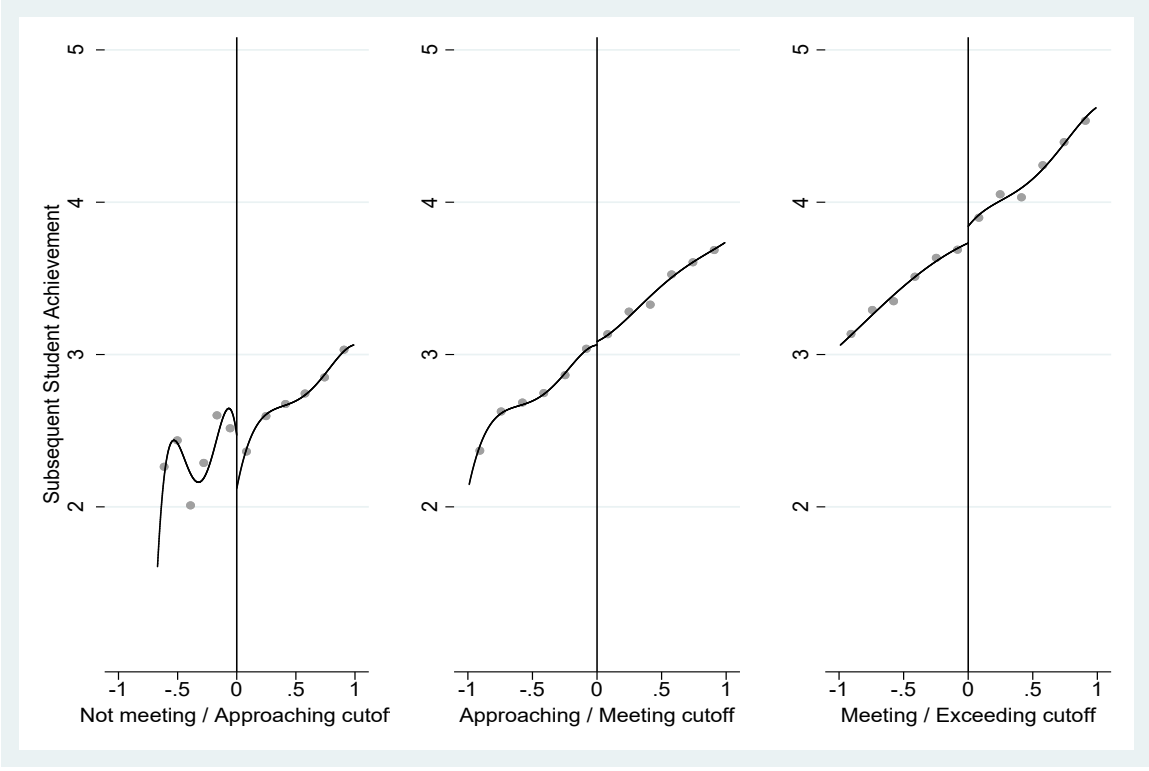


Figure 2 Distribution of student achievement scores and grades

RD results for overall performance improvement

To investigate the effect of past performance levels on overall performance, this paper employs an RD model to estimate the causal effect. Figure 3 presents the plots of the RD model, which illustrates whether there are changes at each cutoff point. As Figure 3 shows, there is the largest discontinuity at the cutoff between 'not meeting' and 'approaching target' grade. The graphs indicate that there are relatively small gaps at the cutoff between 'approaching' and 'meeting target' grade and between 'meeting' and 'exceeding target' grade.



Note: The dots indicate local averages of standardized scores for positive rates. Polynomial fit order of 4

Figure 3 RD plots for subsequent performance outputs

The results of RD models support the results from Figure 3. Consistent across all bandwidths and polynomial forms, schools with 'not meeting target' grade are likely to achieve higher performance scores than schools with 'approaching target' grade in the following year. When schools received the lowest grade, which is the 'not meeting target' grade, these schools are likely to have approximately 0.45 higher performance scores in the subsequent year compared to the schools of 'approaching target' grade. Considering that the distance to the next performance level is 1.00 and most of schools regarded as 'not meeting target' schools have scores above 1.50, the subsequent performance of these schools is above or close to 'approaching target' grade, which is the performance signal that indicates acceptable performance. On the other hand, there are no significant changes at other cutoffs. 'Approaching target' grade and 'meeting target' grade would not play a significant role for schools to improve their overall performance. Hence, the evidence supports hypothesis 1, which implies that low-performing schools improve their performance in a subsequent year.

Table 5 RD Regression Results for subsequent performance outputs

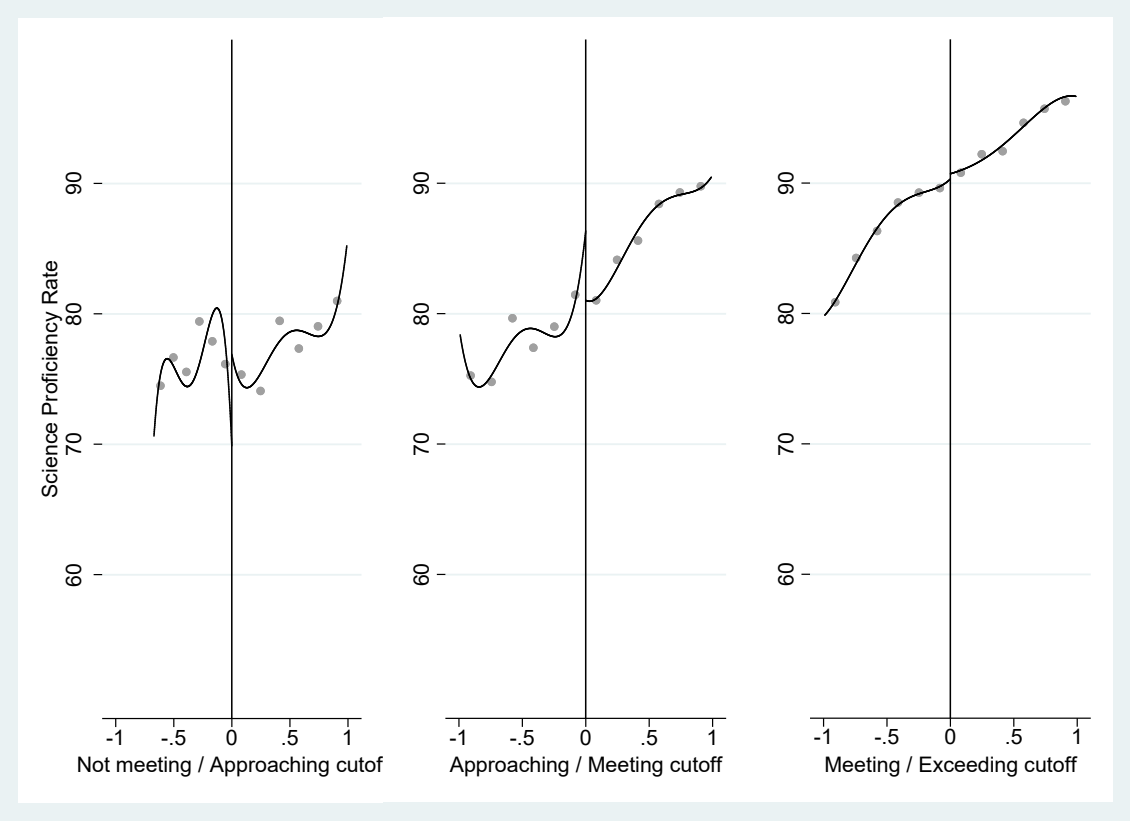
Dependent variables	(1) MSE-optimal bandwidth (mserd) Linear	(2) MSE-optimal bandwidth (msesum) Linear	(3) CER-optimal bandwidth (cerrd) Linear	(4) MSE-optimal bandwidth (mserd) Quadratic
Subsequent Student Achievement				
RD estimator	-0.45***	-0.45**	-0.49***	-0.46**
(Not meeting/Approaching)	[-0.83, -0.12]	[-0.82, -0.1]	[-0.86, -0.13]	[-0.91, -0.03]
Effective N	163	146	121	219
RD estimator	-0.01	-0.01	-0.02	-0.02
(Approaching/Meeting)	[-0.17, 0.12]	[-0.17, 0.14]	[-0.19, 0.13]	[-0.2, 0.17]
Effective N	1392	1294	1019	1653
RD estimator	0.09	0.08	0.09	0.10

(Meeting/Exceeding)	[-0.09, 0.23]	[-0.09, 0.23]	[-0.09, 0.26]	[-0.13, 0.34]
Effective N	846	888	603	818

Note: Inference based on robust bias-corrected confidence intervals that are presented in brackets. All models include year fixed effects and covariates. Standard errors clustered at the school level. * p < 0.1 , ** p < 0.05, *** p < 0.01

RD results for low-stakes goal

Similar to the format in RD results for the overall performance, Figure 3 shows the RD plots based on the low-stakes goal, which is the proficiency rate of Science in this study. Different from the results for overall performance, the proficiency rate of Science markedly dropped near the cutoff between 'not meeting target' and 'approaching target' grades. At the cutoff between 'approaching target' and 'meeting target' grade, the values in 'approaching target' grade spiked near the cutoff point while the values smoothly continues over the cutoff between 'meeting target' and 'exceeding target' grade. The fact that there seems to be a large change at the cutoff between 'approaching target' and 'meeting target' grade let us to check whether it is also statistically significant in the rigorous estimation.



Note: The dots indicate local averages of standardized scores for positive rates. Polynomial fit order of 4

Figure 3 RD plots for the low-stakes goal

Table 5 shows the results with statistical inference. Different from the results in the plots in Figure 3, the results of RD model present that there are no significant discontinuities at the cutoff between 'approaching target' and 'meeting target' grades. The changes of the Science proficiency rate remained as approximately 1.6pp at the cutoff point. There is also no statistical significance at the cutoff between 'meeting target' and 'exceeding target' as we checked in the RD plot. On the other hand, these results show significant changes at the cutoff between 'not meeting target' and 'approaching target' grade. When schools receive 'approaching target' grade, which is relative higher than "not meeting target", they are likely to achieve higher Science proficiency rate. The effect sizes range from 6.2pp to 10.5

pp, depending on the bandwidth and polynomial forms. This means that poor performing schools achieved significantly lower proficiency rate in Science in the subsequent year compared to relatively high performing schools. This could be caused by low-performing schools redirecting resources to high-stakes goals and ignore low-stake goals. Overall, evidence from the results for the low-stakes goal supports Hypothesis 2 that low performing schools have less effort on the goal that is not included in performance metrics.

Table 5 RD Regression results for the low-stakes goal

Dependent variables	(1) MSE-optimal bandwidth (mserd) Linear	(2) MSE-optimal bandwidth (msesum) Linear	(3) CER-optimal bandwidth (cerrd) Linear	(4) MSE-optimal bandwidth (mserd) Quadratic
Science Proficiency Rate				
RD estimator	6.20**	7.34**	10.15**	10.5**
(Not meeting/Approaching)	[0.52, 20.33]	[0.6, 21.27]	[2.69, 22.99]	[1.09, 24.92]
Effective N	115	106	84	156
RD estimator	-1.29	-1.61	-1.74	-1.82
(Approaching/Meeting)	[-4.87, 1.69]	[-5.3, 1.58]	[-5.44, 1.61]	[-6.05, 1.99]
Effective N	1014	825	690	1295
RD estimator	-0.31	-0.35	-0.46	-0.38
(Meeting/Exceeding)	[-2.51, 1.93]	[-2.61, 1.9]	[-2.84, 1.93]	[-3.23, 2.06]
Effective N	1079	1027	762	1360

Note: Inference based on robust bias-corrected confidence intervals that are presented in brackets. All models include year fixed effects and covariates. Standard errors clustered at the school level. * p < 0.1 , ** p < 0.05, *** p < 0.01

Robustness Check: RD model with Expanded Dataset

Although the main RD models confirm that low-performing organizations improve overall performance while they have less effort for the low-stakes goal, the number of observations can threaten the statistical power of the main model. To overcome this concern, this study

utilized the dataset of middle schools. As mentioned earlier, performance metrics for middle schools include the indicator for Social Studies and Science. However, the weight for those indicators is very low in the evaluation metrics. The indicators for Social Studies and Science take 1% each of the overall student achievement grade. Based on this fact, this study made the assumption that the weight of those indicators is not likely to be significant for decision-makers in middle schools when they manage the student achievement ratings.

The results in Table 6 support the main RD model and show the consistent significances at the cutoff between 'not meeting target' and 'approaching target' grades. Compared to the schools with 'approaching target' grade, the schools with 'not meeting target' grade are likely to achieve higher student achievement scores in the subsequent year, and lower Science proficiency rate in the subsequent year. The effect sizes are similar to the effect sizes in the main models. In comparison to the schools with 'approaching target' grade, schools with 'not meeting target' grade are likely to have approximately 0.46 higher student achievement scores while they are likely to achieve about 10 pp lower proficiency rate of Science. This evidence also supports Hypotheses 1 and 2.

Table 6 RD regression results with Dataset Adding Middle Schools

Dependent variables	(1) MSE-optimal bandwidth (mserd) Linear	(2) MSE-optimal bandwidth (msesum) Linear	(3) CER-optimal bandwidth (cerrd) Linear	(4) MSE-optimal bandwidth (mserd) Quadratic
Subsequent Student Achievement				
RD estimator	-0.43**	-0.46**	-0.58***	-0.39*
(Not meeting/Approaching)	[-0.75, -0.01]	[-0.8, -0.05]	[-0.93, -0.17]	[-0.78, 0.06]
Effective N	158	138	107	258
RD estimator	0.04	0.04	0.05	0.04
(Approaching/Meeting)	[-0.08, 0.17]	[-0.09, 0.17]	[-0.09, 0.18]	[-0.11, 0.21]

Effective N	1788	1652	1231	2101
RD estimator	0.06	0.06	0.07	0.08
(Meeting/Exceeding)	[-0.06, 0.16]	[-0.06, 0.16]	[-0.05, 0.18]	[-0.08, 0.24]
Effective N	1492	1492	1065	1242
Science Proficiency Rate				
RD estimator	9.94**	9.96**	15.71***	13.2**
(Not meeting/Approaching)	[1.99, 26.15]	[1.71, 26.36]	[7.04, 30.13]	[2.28, 29.46]
Effective N	117	117	88	213
RD estimator	0.46	0.43	-0.71	-3.72
(Approaching/Meeting)	[-5.11, 6.2]	[-5.46, 6.16]	[-6.74, 5.42]	[-13.47, 3.63]
Effective N	2104	1994	1506	1603
RD estimator	-3.91	-3.55	-3.31	-4.24
(Meeting/Exceeding)	[-12.02, 2.12]	[-11.57, 1.68]	[-11.33, 3.71]	[-12.08, 1.89]
Effective N	937	1172	666	1775

Note: Inference based on robust bias-corrected confidence intervals that are presented in brackets. All models include year fixed effects and covariates. Standard errors clustered at the school level. * p < 0.1, ** p < 0.05, *** p < 0.01

Discussion

This study investigated in how past performance signals influence organizational decision making in multiple goals environments. While previous literature sheds light on goal prioritization in response to performance shortfalls (e.g. Holm, 2018; Nielsen, 2014), this study focused on the unintended effects from the goal prioritization: effort substitution. Using a quasi-experimental design this paper shows that, in response to performance shortfalls, public schools have more interests in high-stakes goals. Results showed that at the cutoff between 'not meeting target' and 'approaching target', schools in the 'not meeting target' grade are likely to improve overall performance while they are less likely to put effort towards a low-stakes goal in performance evaluation. However, other cutoffs, such as between 'approaching target' and 'meeting target' and between 'meeting target' and 'exceeding target', do not show significant discontinuities in terms of the overall performance and outcomes of low-stakes goal.

Accordingly, the results confirm that accountability pressure may explain the behaviors of low-performing organizations regarding goal prioritization and effort substitution. Low performance signals pressure managers to reassign resources and attentions toward high-stakes goals that have greater potential to contribute to performance evaluation outcomes. It therefore positively influences the improvement of performance outcomes. However, this improvement was at the expense of low-stakes goals. Because low-performing organizations intensively focused on high-stakes goals, they have paid relatively less attention on low-stakes goals, which is not closely related to performance evaluations. Hence, the performance pressure may be the reason why effort substitution occurred.

In addition to the relationship between performance shortfalls, performance improvement, and effort substitution, the results of this study allow us to discuss about the accountability mechanisms. Previous literature attributes effort substitution to the rewards and sanctions associated with accountability mechanisms, which provides rewards for high-performers and sanctions for low-performers. This is because prior studies view that accountability pressure could amplify the risk of effort substitution (Han & Wang, 2022; Huizinga & de Bree, 2021). The pressure from the accountability to their performance may lead them to engage in effort substitution. As noted earlier, reward and sanction accountability system often entails benefits and constraints on autonomy; high performers are allowed to use more financial and human resources while low performers have constraints to access them (i.e. Hong, 2019; Min et al., 2021; Wang & Yeung, 2019). Considering that the sanctions are more severe when organizations consecutively receive poor performance grade, such as managers losing job and the closure of organizations, low-

performed organizations have more accountability pressure than high-performed organizations.

Nevertheless, what is unknown in the previous literature is whether low performing organizations could engage less in effort substitution in the naming-and-shaming accountability structure, whose consequences do not include rewards or sanctions. Researchers believe that the failure of public organizations to achieve performance targets drive them to prioritize goals included in performance assessment (Bohte & Meier, 2000b). If sanctions from low performance are less consequential in the naming and shaming accountability structure than they are in the reward-and-sanction accountability structure, it may drive low-performing organizations to care the goals that are essential but are not included in performance evaluation (Mikkelsen et al., 2022). This could alleviate effort substitution. While previous literature suggested several prescriptions towards effort substitution such as amending organization norms (Bohte & Meier, 2000b), reviewing performance measures constantly (Barnow & Heinrich, 2010), enhancing managerial autonomy (Han & Wang, 2022), and strengthening oversight (Gao, 2021), there is little discussion on the different kind of accountability structure.

With those questions remaining from previous literature, the result of this study provides evidence that naming-and-shaming accountability system may work for performance improvement, but it does not remedy the effort substitution. The institutional background of this study is based on the naming-and-shaming accountability system. During the study periods, NYCDOE published school performance information on their website and the public media, instead of rewarding for high performers and sanctioning for low performers. The pressure for low performing schools mainly came from naming-and-

shaming by the media and the public. Hence, this study could add to the evidence that naming-and-shaming accountability system could still incentivize public managers to improve organizational performance. This study's result can contribute to public management research because of the growing body of scholarship that is concerned about negative effects of sanctions. Although reward-and-sanction accountability would motivate public managers to achieve better performance outcomes (Heinrich, 2002), this ratcheted-up accountability entails negative effects on the organizational climate. For example, reward and sanction followed by performance outcomes may induce competitive culture and deteriorate trust and collaboration in colleagues, and reduce intrinsic motivation of employees (Jacobsen & Bøgh Andersen, 2017; Wang & Kim, 2022). In this sense, this study provides empirical evidence that improvement of performance outcome is possible with the accountability system without sanctions, through the transparency in information to the public.

The results of this study showed that underperforming organizations neglect the low-stakes goal in the process of increasing efficiency and effectiveness similar to the previous literature which explained goal displacement based on performance management (e.g. Bohte & Meier, 2000; Huizinga & de Bree, 2021). However, the naming-and-shaming accountability may not be a solution to mitigate effort substitution. This finding points to a different result from the views presented in previous literature on accountability, which predicted the relaxed accountability pressure as a potential solution to reduce the cheating and gaming behaviors on performance management (Heinrich & Marschke, 2010; Mikkelsen et al., 2022; Muller, 2018). This could be because public managers recognize

reputation damages from naming-and-shaming accountability as significant sanctions for them.

Conclusion

This paper examined the managerial behaviors that managers prioritize the goals accounted towards performance evaluation while they pay less attention to the goals underrepresented. However, the way of decreasing effort substitution could be complex. Considering that previous literature on performance management discussed overrepresentation from a small number of performance indicators (Van Thiel & Leeuw, 2002), it led to the conclusive recommendations to add more goals in performance evaluations and increase the weight of goals that are underrepresented. However, adding the number of performance indicators may not be the panacea to reduce effort substitution in low performing organizations. First, as we discussed earlier, public managers often prioritize goals because of efficiency and bounded rationality (Holm, 2018; Hong et al., 2020b). Although policy makers design multiple performance indicators for low-stakes goals, it would not be effective if there are already numerous indicators that managers are required to achieve simultaneously. Second, too many performance indicators may increase goal ambiguity, and therefore does not suit the purpose of performance management. Performance management is introduced to resolve goal ambiguity (Huizinga & de Bree, 2021). This means that performance management has a mandate to specify the core goals of organizations and measure the extent to which organizations achieve their goals. Overall, simply increasing the number of performance indicators may not be the solution.

Potential solutions of effort substitution could be enhanced managerial autonomy and internal learning with the effort to balance external accountability. Managerial autonomy may reduce the gaming behaviors, including effort substitution (Han & Wang, 2022). Autonomy enables public managers to take a flexible approach to respond to the diverse needs of citizens, and it motivates them to have ownership, and make necessary and timely adjustments. Additionally, the introduction of internal learning process, which "combines political-bureaucratic performance goals with extensive professional involvement in the development and interpretation of goals (Jakobsen et al., 2018: 128)," would be a positive drive to mitigate the effort substitution. Lastly, periodical involvement of professionals would enable performance management to reflect task complexity and emerging issues in practices through facilitating the discussion in terms of performance metrics among regulators, professionals, and practitioners.

Despite the rigorous empirical design with RDD that shed light on the relationship between performance shortfalls, performance improvement, and effort substitution, this paper has limitations. The first limitation is generalizability. Since this study utilized the New York City public school data, it is understandable that people may be skeptical whether the results and lessons could be applicable to other public organizations. However, over the years, education has been emerging sector to test the theory and give lesson in performance management (Han & Wang, 2022; Holm, 2018; Nielsen, 2014a; Wang & Kim, 2022; Wang & Yeung, 2019). Furthermore, education and other public sectors share common characteristics; employees are managed by a merit system, public fund is the main resource of funding to operate, and managers are held accountable to citizens. Secondly, it could be limiting that a single indicator is used to measure effort substitution. This study

utilized the proficiency rate in Science to measure effort substitution. Whereas using other indicators such as Social Studies would have strengthened this study, there was no data for proficiency rate of Social Studies during the study periods.

Despite the limitations of this study, there are substantial outcomes of this study that contribute to both theory and practice. First, it adds to the evidence for the relationship between performance management and effort substitution. Specifically, this study focused on the understudied area of effort substitution through highlighting low performing organizations. Previous literature has paid more attention to the performance regimes such as managing results and task complexity (Benaine & Kroll, 2020; Han & Wang, 2022). However, this study focused on performance signals and how organizational behaviors diverge based on those signals. Second, it deepens our understanding of naming-and-shaming accountability system. This paper provides evidence that low performing organizations are still motivated to improve future performance without rewards or sanctions and that they also engage in effort substitution. Considering that empirical evidence that naming-and-shaming accountability is based on is severely limited, the lessons of this study could facilitate the discussion for an alternative accountability system. The evidence indicates that accountability without sanctions would be enough to make public managers to be held accountable if performance information is transparent to the public. Policy makers and practitioners may consider introducing naming-and-shaming instead of rewards-and-sanctions accountability mechanism as rewards-and-sanction often raise side effects on organizational climates and annual budget forecast (Wang & Kim, 2022). To reduce effort substitution of low performing organization, however, it may be

necessary to have additional factors including managerial autonomy and internal learning (Han & Wang, 2022; M. L. Jakobsen et al., 2018).

References

Andersson, T., & Liff, R. (2012). Multiprofessional Cooperation and Accountability Pressures: Consequences of a post-new public management concept in a new public management context. *Public Management Review*, *14*(6), 835–855. https://doi.org/10.1080/14719037.2011.650053

Argote, L., & Greve, H. R. (2007). *A Behavioral Theory of the Firm* —40 Years and Counting: Introduction and Impact. *Organization Science*, *18*(3), 337–349. https://doi.org/10.1287/orsc.1070.0280

Barnow, B. S., & Heinrich, C. J. (2010). One Standard Fits All? The Pros and Cons of Performance Standard Adjustments. *Public Administration Review*, *70*(1), 60–71. https://doi.org/10.1111/j.1540-6210.2009.02111.x

Benaine, S. L., & Kroll, A. (2020). Explaining effort substitution in performance systems: The role of task demands and mission orientation. *Public Management Review*, *22*(6), 813–835. https://doi.org/10.1080/14719037.2019.1604794

Bevan, G., & Wilson, D. (2013). Does 'naming and shaming' work for schools and hospitals? Lessons from natural experiments following devolution in England and Wales. *Public Money & Management*, *33*(4), 245–252. https://doi.org/10.1080/09540962.2013.799801

Bohte, J., & Meier, K. J. (2000). Goal Displacement: Assessing the Motivation for Organizational Cheating. *Public Administration Review*, *60*(2), 173–182. https://doi.org/10.1111/0033-3352.00075

Boyne, G. A., James, O., John, P., & Petrovsky, N. (2009). Democracy and Government Performance: Holding Incumbents Accountable in English Local Governments. *The Journal of Politics*, *71*(4), 1273–1284. https://doi.org/10.1017/S0022381609990089

Boyne, G. A., Meier, K. J., Meier, K. J., O'Toole Jr, L. J., Walker, R. M., & others. (2006). *Public service performance: Perspectives on measurement and management*. Cambridge University Press.

Carpenter, D. P., & Krause, G. A. (2012). Reputation and Public Administration. *Public Administration Review*, *72*(1), 26–32. https://doi.org/10.1111/j.1540-6210.2011.02506.x

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). A Practical Introduction to Regression Discontinuity Designs: Foundations. *Elements in Quantitative and Computational Methods for the Social Sciences*. https://doi.org/10.1017/9781108684606

Chen, S., & Jia, K. (2021). How local governments prioritize multiple conflicting goals: Beyond the sole-goal perspective. *Public Administration*.

Christensen, J., Dahlmann, C. M., Mathiasen, A. H., Moynihan, D. P., & Petersen, N. B. G. (2018). How Do Elected Officials Evaluate Performance? Goal Preferences, Governance Preferences, and the Process of Goal Reprioritization. *Journal of Public Administration Research and Theory*, *28*(2), 197–211. https://doi.org/10.1093/jopart/muy001

Cyert, R. M., March, J. G., & others. (1963). *A behavioral theory of the firm* (Vol. 2). Englewood Cliffs, NJ.

Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, *35*(2), 252–279.

Downe, J., Grace, C., Martin, S., & Nutley, S. (2010). Theories Of Public Service Improvement: A comparative analysis of local performance assessment frameworks. *Public Management Review*, *12*(5), 663–678. https://doi.org/10.1080/14719031003633201

Gao, J. (2021). Mitigating Pernicious Gaming in Performance Management in China: Dilemmas, Strategies and Challenges. *Public Performance & Management Review*, *44*(2), 321–351. https://doi.org/10.1080/15309576.2020.1775662

Gilad, S. (2015). POLITICAL PRESSURES, ORGANIZATIONAL IDENTITY, AND ATTENTION TO TASKS: ILLUSTRATIONS FROM PRE-CRISIS FINANCIAL REGULATION: REGULATORY ALLOCATION OF ATTENTION. *Public Administration*, *93*(3), 593–608. https://doi.org/10.1111/padm.12155

Girth, A. M. (2014). A Closer Look at Contract Accountability: Exploring the Determinants of Sanctions for Unsatisfactory Contract Performance. *Journal of Public Administration Research and Theory*, *24*(2), 317–348. https://doi.org/10.1093/jopart/mus033

Glasman, N. S. (1984). Student achievement and the school principal. *Educational Evaluation and Policy Analysis*, *6*(3), 283–296.

Greve, H. (2003). *Organizational Learning from Performance Feedback*. Cambridge.

Han, X., & Wang, W. (2022). Does granting managerial autonomy in exchange for accountability mitigate gaming? *Public Administration Review*, puar.13564. https://doi.org/10.1111/puar.13564

Heinrich, C. J. (2002). Outcomes-Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness. *Public Administration Review*, *62*(6), 712–725. https://doi.org/10.1111/1540-6210.00253

Heinrich, C. J. (2012). How Credible Is the Evidence, and Does It Matter? An Analysis of the Program Assessment Rating Tool. *Public Administration Review*, *72*(1), 123–134. https://doi.org/10.1111/j.1540-6210.2011.02490.x

Heinrich, C. J., & Marschke, G. (2010). Incentives and their dynamics in public sector performance management systems. *Journal of Policy Analysis and Management*, *29*(1), 183–208. https://doi.org/10.1002/pam.20484

Holm, J. M. (2018). Successful Problem Solvers? Managerial Performance Information Use to Improve Low Organizational Performance. *Journal of Public Administration Research and Theory*, *28*(3), 303–320. https://doi.org/10.1093/jopart/muy017

Hong, S. (2019). A Behavioral Model of Public Organizations: Bounded Rationality, Performance Feedback, and Negativity Bias. *Journal of Public Administration Research and Theory*, *29*(1), 1–17. https://doi.org/10.1093/jopart/muy048

Hong, S., Kim, S. H., & Son, J. (2020). Bounded rationality, blame avoidance, and political accountability: How performance information influences management quality. *Public Management Review*, *22*(8), 1240–1263. https://doi.org/10.1080/14719037.2019.1630138

Huizinga, K., & de Bree, M. (2021). Exploring the Risk of Goal Displacement in Regulatory Enforcement Agencies: A Goal-Ambiguity Approach. *Public Performance & Management Review*, *44*(4), 868–898. https://doi.org/10.1080/15309576.2021.1881801

Jacobsen, C. B., & Bøgh Andersen, L. (2017). Leading public service organizations: How to obtain high employee self-efficacy and organizational performance. *Public Management Review*, *19*(2), 253–273. https://doi.org/10.1080/14719037.2016.1153705

Jakobsen, M. L., Baekgaard, M., Moynihan, D. P., & van Loon, N. (2018). Making Sense of Performance Regimes: Rebalancing External Accountability and Internal Learning. *Perspectives on Public Management and Governance*, *1*(2), 127–141. https://doi.org/10.1093/ppmgov/gvx001

Ji, S. (2022). Role of Managerial and Political Accountability in Improving Performance of Local Governments. *Public Performance & Management Review*, *45*(3), 499–531. https://doi.org/10.1080/15309576.2022.2062608

Kelman, S., & Friedman, J. N. (2009). Performance Improvement and Performance Dysfunction: An Empirical Examination of Distortionary Impacts of the Emergency Room Wait-Time Target in the English National Health Service. *Journal of Public Administration Research and Theory*, *19*(4), 917–946. https://doi.org/10.1093/jopart/mun028

Koning, P., & Heinrich, C. J. (2013). Cream-skimming, parking and other intended and unintended effects of high-powered, performance-based contracts. *Journal of Policy Analysis and Management*, *32*(3), 461–483.

Kuusela, P., Keil, T., & Maula, M. (2017). Driven by aspirations, but in what direction? Performance shortfalls, slack resources, and resource-consuming vs. resource-freeing organizational change: Driven by Aspirations, but in What Direction. *Strategic Management Journal*, *38*(5), 1101–1120. https://doi.org/10.1002/smj.2544

Lee, G., & Jimenez, B. S. (2011). Does Performance Management Affect Job Turnover Intention in the Federal Government? *The American Review of Public Administration*, *41*(2), 168–184. https://doi.org/10.1177/0275074010368991

Lu, Y., Yang, K., & Thomas, M. B. (2021). Designing Performance Systems in Anticipation of Unintended Consequences: Experiences and Lessons From the Compstat-Based Performance Regime in NYPD. *Administration & Society*, *53*(6), 907–936. https://doi.org/10.1177/0095399720976532

March, J. G., & Simon, H. A. (1958). *Cognitive Limits on Rationality.*

Meier, K. J., Favero, N., & Zhu, L. (2015). Performance Gaps and Managerial Decisions: A Bayesian Decision Theory of Managerial Action. *Journal of Public Administration Research and Theory*, *25*(4), 1221–1246. https://doi.org/10.1093/jopart/muu054

Meier, K. J., & O'Toole, L. J. (2003). Public Management and Educational Performance: The Impact of Managerial Networking. *Public Administration Review*, *63*(6), 689–699. https://doi.org/10.1111/1540-6210.00332

Mikkelsen, M. F., Petersen, N. B. G., & Bjørnholt, B. (2022). Broadcasting good news and learning from bad news: Experimental evidence on public managers' performance information use. *Public Administration*, *100*(3), 759–777. https://doi.org/10.1111/padm.12762

Min, B. H., Oh, Y., & Brower, R. S. (2021). The Effects of Diverse Feedback Dynamics on Performance Improvement: A Typology of Performance Feedback Signals. *Administration & Society*, *53*(1), 123–157. https://doi.org/10.1177/0095399720933822

Moynihan, D. P., Fernandez, S., Kim, S., LeRoux, K. M., Piotrowski, S. J., Wright, B. E., & Yang, K. (2011). Performance Regimes Amidst Governance Complexity. *Journal of Public Administration Research and Theory*, *21*(Supplement 1), i141–i155. https://doi.org/10.1093/jopart/muq059

Muller, J. Z. (2018). The tyranny of metrics. In *The Tyranny of Metrics*. Princeton University Press.

Nettles, S. M., & Herrington, C. (2007). Revisiting the importance of the direct effects of school leadership on student achievement: The implications for school improvement policy. *Peabody Journal of Education*, *82*(4), 724–736.

Nicholson-Crotty et al. (2017). *Performance and Management in the Public Sector: Testing a Model of Relative Risk Aversion:* 12.

Nielsen, P. A. (2014). LEARNING FROM PERFORMANCE FEEDBACK: PERFORMANCE INFORMATION, ASPIRATION LEVELS, AND MANAGERIAL PRIORITIES: Learning from performance feedback. *Public Administration*, *92*(1), 142–160. https://doi.org/10.1111/padm.12050

Posen, H. E., Keil, T., Kim, S., & Meissner, F. D. (2018). Renewing Research on Problemistic Search—A Review and Research Agenda. *Academy of Management Annals*, *12*(1), 208–251. https://doi.org/10.5465/annals.2016.0018

Radin, B. A. (2006). *Challenging the performance movement: Accountability, complexity, and democratic values*. Georgetown University Press.

Rutherford, A., & Meier, K. J. (2015). Managerial goals in a performance-driven system: Theory and empirical tests in higher education. *Public Administration*, *93*(1), 17–33.

Salge, T. O. (2011). A Behavioral Model of Innovative Search: Evidence from Public Hospital Services. *Journal of Public Administration Research and Theory*, *21*(1), 181–210. https://doi.org/10.1093/jopart/muq017

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129.

Simon, H. A. (1966). Theories of decision-making in economics and behavioural science. In *Surveys of economic theory* (pp. 1–28). Springer.

Soehner, D., & Ryan, T. (2011). *The Interdependence of Principal School Leadership and Student Achievement*. *5*(3).

Sohn, H., & Bae, K. B. (2018). Do Public Employees "Game" Performance Budgeting Systems? Evidence From the Program Assessment Rating Tool in Korea. *The American Review of Public Administration*, *48*(5), 458–475. https://doi.org/10.1177/0275074016689322

Taylor, J. (2021). Public Officials' Gaming of Performance Measures and Targets: The Nexus between Motivation and Opportunity. *Public Performance & Management Review*, *44*(2), 272–293. https://doi.org/10.1080/15309576.2020.1744454

Van Thiel, S., & Leeuw, F. L. (2002). The performance paradox in the public sector. *Public Performance & Management Review*, *25*(3), 267–281.

Walker, R. M., Damanpour, F., & Devece, C. A. (2011). Management Innovation and Organizational Performance: The Mediating Effect of Performance Management. *Journal of Public Administration Research and Theory*, *21*(2), 367–386. https://doi.org/10.1093/jopart/muq043

Wang, W. (2022). How does performance management affect social equity? Evidence from New York City public schools. *Public Administration Review*, puar.13590. https://doi.org/10.1111/puar.13590

Wang, W., & Kim, T. K. (2022). Examining the effects of a performance management reform on employee attitudes and organizational climate. *Public Management Review*, 1–23. https://doi.org/10.1080/14719037.2022.2026095

Wang, W., & Yeung, R. (2019). Testing the Effectiveness of "Managing for Results": Evidence from an Education Policy Innovation in New York City. *Journal of Public Administration Research and Theory*, *29*(1), 84–100. https://doi.org/10.1093/jopart/muy043

Zhu, L., & Rutherford, A. (2019). Managing the Gaps: How Performance Gaps Shape Managerial Decision Making. *Public Performance & Management Review*, *42*(5), 1029–1061. https://doi.org/10.1080/15309576.2019.1568886

Chapter 3: Performance Management and Employee Behaviors: How Do Organizational Performance Signals Affect Employee Fraud?

Introduction

While performance management reduces goal ambiguity and enables public organizations to systematically manage their goals, it can also lead to organizational cheating, which deliberately manipulates output levels to achieve the desirable performance scores or ratings (Bohte & Meier, 2000b; Heckman et al., 2002; Koning & Heinrich, 2013). In order to achieve desired performance outcomes, public managers could engage in various strategic behaviors which may be adopted in conjunction or exclusively: "data manipulation", which distorts the performance outputs (Chen, 2023; Kroll & Vogel, 2021); "effort substitution", which is devoted only to high-stakes performance indicators (Benaine & Kroll, 2020; Kelman & Friedman, 2009); "cream-skimming", which produces public services for recipients who are easier and less costly to serve (Heckman et al., 2002; Hood, 2007; Koning & Heinrich, 2013). These strategic behaviors breach social norms and institutions, such as providing distorted information to citizens and damaging public trust (Heinrich & Marschke, 2010).

Despite the contributions from previous literature, the attention has mostly been focused on cheating at the organization level. There is still little discussion on the following question: How do performance shortfalls affect employee fraud? Employee fraud is different from organizational cheating in that it is the employee behaviors to gain financial

advantages for personal enrichment by misusing organizational resources (Case, 2000; Greenberg, 1990). Employee fraud mainly benefit employee themselves rather than the organizations. It distorts public service production and also damages organizational climate (Fox et al., 2001; Marcus & Schuler, 2004). Therefore, it is essential to investigate occurrence of employee fraud in the public sector. However, rarely have researchers examined employee fraud as a plausible behavior in response to performance shortfalls. On the other hand, the literature on performance feedback have mostly focused on legitimate forms of organizational changes in response to low performance, such as the re-establishment of an organization's goals (Holm, 2018; Nielsen, 2014b; Salge, 2011b) and organizational innovation (Meier et al., 2015; Nicholson-Crotty et al., 2017). They were also more interested in negativity bias (Hong, 2019b; Hong & Kim, 2019) and problemistic search (Van der Voet, 2022; van der Voet, 2023).

This paper bridges performance management literature and employee fraud literature by examining whether performance shortfalls affect employee fraud. There are three rationales that this study predicts the significant relationship between low performance signals and employee fraud. First, managers of low performing organizations are under accountability pressure, which may cause them to engage in organizational cheating. These behaviors deteriorate ethical environment in their organizations, which lower the cost of engaging in employee fraud, change one's understanding of social norms, and increase the cost to keep their moral self-image (Gino et al., 2009; Rickman & Witt, 2007). That is, low-performing organizations where organizational cheating is highly likely to occur may increase the risk of occurrence of employee fraud. Second, the lack of financial rewards to employees in low performing organizations may influence employee

fraud. When it comes to accountability systems that impose financial penalties as a response to organizational performance shortfalls, employees' perception is negatively impacted; low performance signals could raise employees' feeling of inequity and inequality on their financial benefits compared to other public organizations (Goh & Kong, 2018; Greenberg, 2002; Peters & Maniam, 2016). When employees perceived that they are unjustifiably treated in financial penalty compared to other public organizations, the probability of employee fraud may increase.

To test this relationship, this study links data from performance evaluation and integrity assessment for South Korean government corporations. Regression discontinuity design (RDD) was employed to estimate the impact of performance shortfalls. The annual performance evaluation system in Korea measures the performance of government corporations, both quantitatively and qualitatively. Grades are clear-cut, and all observations follow the cutoff rule. Grades are a criterion for rewards and sanctions. On the other hand, in Korea there is the annual integrity assessment which is a separate system to evaluate ethical climates in government agencies, including local governments and government corporations. The integrity assessment scores the degree of violating expense policies through employee and citizen surveys. The results show that employees in low-performing public corporates are more likely to abuse operating expenses, benefit allowance, and project expenses than employees in relatively high-performing organizations.

This study makes several contributions on public administration literature. First, it offers a novel perspective on the relationship between performance outcomes and counterproductive work behaviors. While prior studies on performance management paid

more attention to organizational cheating related to performance growth, this paper provides evidence that employees in low-performing organizations may have increased likelihood to engage in employee fraud that harms efficiency, and eventually, organizational performance outcomes. Thus, it fosters research on the determinant of counterproductive work behaviors. Second, it could contribute to research on how accountability systems affect organizational climate and employee ethical behaviors. Literature on accountability system has often focused on unintended effects from sanctioning of low-performing organizations. Different from the expectations that sanctions pressure low-performing organizations to improve their performance outcomes, sanctions could cause the deterioration of working climates and intrinsic motivations of employees (Jacobsen & Bøgh Andersen, 2017; Wang & Kim, 2022). The evidence provided by this paper deepens our understanding of the unintended consequences of accountability system.

Literature Review

Organization Cheating and Performance Management

Previous literature on performance management has paid significant attention to the relationship between performance management and organizational cheating. Organizational cheating related to performance management comes from strategic behaviors to maximize organizational performance ratings (Lynn, 1996). There are several reasons why public organizations engage in organizational cheating related to performance management. The first plausible reason is accountability pressure. In tandem with the New

Public Management movement, modern governments especially emphasize the responsibilities of performance outcomes to their agencies. For instance, Heckman et al. (2002) suggested a performance management system emphasizing outcomes rather than short-term impacts of programs as the reason for organizational cheating in public organizations. Another reason for public organizations to possibly engage in organizational cheating is due to its benefits distributed to almost everyone within the organization. One trait of New Public Management is its reward, which is shared by employees when organizations achieve high-performance (Moon 2000). As their organization achieves high-performance, managers and employees may receive performance-related pay and reputations that may be helpful for future potential job transitions. This structure may encourage top managers to ask their employees to do unjustified work for one's own and other colleagues' rewards. Kroll and Vogel (2021) also pointed out systemic factors, such as red tape, as a reason for organizational cheating.

Furthermore, prior studies found two types of organizational cheating related to the performance management process. The first type is gaming behaviors in setting performance targets (Bevan & Hood, 2006; Goddard et al., 2000). Public organizations have attempted to set lower performance targets to make them easier to achieve. However, it is not a common type of organizational cheating in performance management these days since decent performance management is not likely to authorize organizations to set their performance targets. Rather, current performance management sets the targets with a formula and in consultation with regulators (Hong, 2019b; Portz, 2021). Most common organizational cheating in performance management occurs in the process of achieving performance targets. These include "data manipulation", which distorts the performance

outputs (Chen, 2023; Kroll & Vogel, 2021); "effort substitution", which is devoted only to high-stakes performance indicators (Benaine & Kroll, 2020; Kelman & Friedman, 2009); and "cream-skimming", which produces public services for favorable and low-severity recipients who are easier and less costly to serve (Heckman et al., 2002; Hood, 2007; Koning & Heinrich, 2013). There is also the "cutting corners" behavior to meet the level of public service production without the consideration of service quality (Bohte & Meier, 2000b). In connection to these strategic behaviors, for example, public schools may concentrate on Math and English Language Art, which are high-stakes indicators of overall performance, while less attention is paid to Science which is often a low-stakes subject (Benaine & Kroll, 2020; Han & Wang, 2022). School teachers may pay great attention to marginal students who are close to the target thresholds (Taylor, 2021). In other cases, municipal governments manipulate the air quality data to meet their performance target (Chen, 2023), and regulatory agencies tend to abuse the authority of inspection to maximize the numbers that measure their work efficiency (Bohte & Meier, 2000b).

Puzzlingly, however, what prior studies have under-researched is the relationship between performance management and employee fraud, which is "the practice of taking organization property for personal use (Greenberg, 2002, p. 985)." It is a common concern in organizations; most organizations aim to prevent occurrence of employee fraud and theft (Case, 2000). Although organizational cheating harms more on public values, such as equality and equity, employee fraud negatively influences both on efficiency and public values (Owusu et al., 2022; Rustiarini et al., 2019). Hence, considering that employee fraud disadvantages other members of the organization, the organizations, and even public

service production (Fox et al., 2001; Marcus & Schuler, 2004), it is essential to investigate the specific factors which influences the occurrence of employee fraud in the public sector.

Employee fraud and Organizational Performance outcomes

Public administration scholars have paid less attention to the relationship between employee fraud and organizational performance. Here organizational performance is measured by level of efficiency. Instead, public administration literature focused more on how ethic management tools, such as codes of conduct, audits, ethical leadership, and ethical climate, affect performance outcomes (Khaltar & Moon, 2020). According to previous literature, those tools are expected to improve organizational performance since they induce employees to be accountable and prevent waste of resources (Adler & Borys, 1996; Burke & Black, 1990). Empirical evidence also supports this expectation. Most studies conclude that performance outcomes have positive relationships with adopting ethic management tools, such as ethical climate (Berman & West, 1997; Menzel, 2015) and ethical leadership (Hassan et al., 2014; Khaltar & Moon, 2020). Only a few studies did not find a significant relationship between ethic management and organizational performance (Burke & Black, 1990; Menzel, 2016).

The relationship between employee fraud and organizational performance was investigated by scholars in business ethics and applied psychology. Evidence from the literature showed a negative impact of employee fraud on organizational performance. Considering the term of 'employee fraud' as employees abusing organizational financial resources for personal use, it imposes hidden costs on organizations and lowered revenue

(Case, 2000; Greenberg, 1990). For example, expense inflation and fictitious expense would lead organizations to increase finance resources spending, and therefore result in lower organizational profits and revenues (Peters & Maniam, 2016). Employee fraud also harms organizational performance by deteriorating reputations and organizational climates, such as employees' commitments (Bhatti et al., 2011), such that it becomes not a worker-friendly workplace and a dishonest organization (Blazovich et al., 2013; Johnson et al., 2014). Those factors lower employee motivation to work, prevent hiring competent employees, and reduce demands of services and products, which are crucial for achieving satisfactory level of organizational performance.

Although some studies in public and business management literature have paid attention to how employee fraud or ethic management affects performance outcomes, interestingly, there has been little discussion about the inverse relationship of how performance outcomes affect employee fraud. Furthermore, the literature on performance feedback concentrated more on other adjustment strategies. For instance, it focused more on how performance signals affect goal reprioritization (Holm, 2018; Nielsen, 2014b; Salge, 2011b), innovation (Meier et al., 2015; Nicholson-Crotty et al., 2017), and alternative searching behaviors (Van der Voet, 2022; van der Voet, 2023). Therefore, in context of performance management, there is a gap to be explored.

Hypothesis: Employee fraud in Low-Performing Organizations

Low performing organizations are more likely to face more accountability pressure compared to organizations with higher performance ratings. Beside organizational cheating,

performance shortfalls could raise the risk of employee fraud in three ways: by damaging the ethical climate in organizations, decreasing financial rewards and threatening self-esteem of employees. First, a low ethical climate in low performing organizations may cause employee fraud. That is, organizational cheating in low performing organizations could lead to employee fraud. Because attempts to fraud, similar to other unethical conducts, are highly influenced by the environment in which it occurs. Ethical climate, such as ethical leadership and ethical culture, plays a critical role in preventing employee fraud (Kuenzi et al., 2019; Thaler & Helmig, 2016). Managers of low performing organizations may engage in cheating behaviors such as data manipulation or cream skimming when they are under performance pressure (Chen, 2023; Han & Wang, 2022). These behaviors deteriorate the ethical environment of their organizations. Employees' understanding of social norms and their readiness to pay greater cost to keep their moral self-image are negatively influenced (Gino et al., 2009). That is, employees may be motivated to justify their theft with other unethical behaviors caused by performance shortfalls (Greenberg, 1997).

Additionally, lack of financial rewards in low-performing organizations could also influence employee fraud. It comes from the high-stakes accountability regimes, which is a dominate form of accountability system in the public sector (Zhu & Rutherford, 2019). Performance shortfalls cause financial sanctions on low performers (Dizon-Ross, 2020; Min et al., 2021; Sohn & Bae, 2018). In this sense, financial rewards and wages are important antecedents of employee fraud (Bhatti et al., 2011; Goh & Kong, 2018; Peters & Maniam, 2016). Specifically, previous literature points out that employee fraud is more likely to be affected by the perception of lack of financial rewards rather than the absolute

level of financial rewards. Here the perception of lack of financial rewards include the feeling of inequity and inequality on their financial benefits, compared to other comparable organizations (Greenberg, 1990, 2002). When employees perceived that they are unjustifiably treated with financial penalty compared to other public organizations, the probability of employee fraud may increase. In addition, performance metrics in the public sector could add to the motivation to engage in employee fraud. Different from the private sector, performance outcomes in the public sectors is not measured by clear-cut indicators; performance measure in the public sector covers multidimensional factors, reflects goal ambiguities, and is often intervened by political factors (Chun, 2005; Julnes & Holzer, 2001; Taylor, 2009). This leads to a lack of consensus on what the valid performance indicators are and raises doubt about the results of performance evaluations (Brignall & Modell, 2000; Kim, 2004). Ambiguities of performance metrics cast doubt over the validity of performance measurement, which further worsen employees feeling of injustice compared with their peers in high-performing organizations. Overall, the gap of financial rewards with weak institutional consensus may raise the feeling of inequality and inequity in financial rewards, and therefore increase the risk of employee fraud in low-performing organizations. This scenario may be encapsulated into following hypothesis:

Hypothesis: Employee fraud is more likely to occur in low-performing organizations than in high-performing organizations

Institutional Background and Empirical Strategy

State-Owned Enterprises of Korea and Performance Evaluation

This paper discusses the employee fraud in response to past performance signals from the case of South Korean state-owned enterprises (hereafter "government corporations") and their performance management systems. These government corporations were established to supply state-owned utilities, including electricity, transportation, water, and energy, and provide public services, such as medical service for veterans (Selin and Lewis 2018; Hong 2019). Moreover, the central government ministries oversee relevant government corporations. Case in point, as noted earlier, the Korean Ministry of Land, Infrastructure, and Transport controls the Incheon International Airport Corporation. Performance evaluations for Korean government corporations were created in the mid-1980's when many governments had designed performance management practices in tandem with the new public management (NPM) movement (Pollit et al. 2001; Hong and Kim 2017; Hong 2019). This evaluation system underwent major reforms in 2007 and has not changed since. The performance evaluation of Korean government corporations is managed by the Korean Ministry of Economy and Finance.

This annual performance evaluation system measures the performance of government corporations both quantitatively and qualitatively. Quantitative performance is evaluated based on government corporation reports submitted to the Korean Ministry of Economy and Finance. Before submission, each report is audited by an external audit firm. These reports include quantitative goals and their achievements. Government corporations set their quantitative goals based on an agreement determined by an independent committee consisting of experts, such as academics, accountants, and lawyers (Hong 2019). This

committee reviews the quantitative reports for any serious errors. Qualitative performance is evaluated by an independent committee. If the role of the expert committee conducting a quantitative performance evaluation is to check for errors, in a qualitative performance evaluation, committee members grade qualitative achievements by visiting the government corporation and interviewing its employees.

There are five performance grades in this evaluation system: A, B, C, D, and E. Grades are assigned through a standardized score. The independent committee creates raw scores by summarizing both quantitative and qualitative performance evaluations, transforming them into standardized scores. For instance, government corporations located above two standard deviation or above from the mean get an A grade; government corporations located between 0 and 1 standard deviation above the mean receive a B grade; government corporations located two standard deviation below the mean get an E grade. These grades are a criterion for distributing performance-related incentives. When agencies get a C, B, or A grade, their employees will receive financial incentives corresponding to 100%, 150%, and 200% of their monthly payroll respectively. However, employees in agencies that get a D or E grade will not get a financial bonus. Furthermore, top managers may be discharged if their agency receives a D or E grade for two consecutive years (Hong 2019). These performance grades are utilized as an index of the top executives' competence. Top executives in government corporations are eager to receive a B grade or above to show their competence and receive promotions (Hong 2019).

Exploring the Generalizability of the Korean Case

The case of Korean government corporations is generalizable to other contexts because many countries have state-owned enterprises and similar performance management systems. Regarding government corporations, this type of public organization is not rare in public management. Many modern governments have government corporations or similar entities to provide public goods; European, Asian, and African countries, even the United States, run these types of public organizations (Bruton et al. 2015). For instance, developed countries, such as England, France, and Germany, and developing countries, such as China, Brazil, Russia, and India, all have government corporations (Kowalski et al. 2013). The United States, whose government prefers privatization (Henig 1989), also has government-owned corporations, such as the United States Postal Service (USPS) and Tennessee Valley Authority (TVA), a federally-owned electric utility corporation. Furthermore, the size of government corporations is not negligible in the global market; government corporations generate approximately 10% of the world's gross domestic product (GDP) and take about 20% of global equity market value (Bruton et al., 2015). In 2011, 204 out of 2,000 of the largest global companies were government corporations (Kowalski et al. 2013).

The organizational structure of these government corporations is also well-matched with trends in the New Pubic Management (NPM) era; this paradigm emphasizes the market mechanisms and flexibility in public organizations (Bertelli 2006a). Some government corporations are hybrid organizations that incorporate both state and private ownership (Inoue et al. 2013). That is, government corporations have different institutional logics; government corporations have a contractual relationship with the government and, also, are not directly linked to legislatures (Bertelli 2006a). Considering the legislature

controls traditional public organizations, the management of government corporations is more market-driven; giving government corporations more flexibility. These traits correspond with the NPM, which is the dominant paradigm in public management (O'flynn 2007; Bryson et al. 2014).

The performance management system for Korean government corporations is similar to those found in other countries. To make sure that government corporations are accountable, governments utilize an auditing mechanism to manage their performances (Bertelli 2006b). For instance, similar to the Korean case, government corporations in the Netherlands and England should submit annual plans, reports, and accounts for performance evaluation. Additionally, performance targets in these agencies are based on agreements by the government corporates, but governments also allow them to use a large amount of discretion to achieve their performance goals (Bertelli 2006a). The Korean performance management system is also similar to some performance regimes in the US , in that they give a cosiderable amount of managerial authority to managers but also require greater accountability. For example, the Empowerment Zone (EZ) undertaken by the New York City Department of Education (NYCDOE) guarantees authority and autonomy of principals but they also may be dismissed should their performance target over the course of 2 years is not met; additionally, the school may be forced to close if the new principal still fails to improve student performance in the following 2 years (Wang and Yeung 2019). These traits are similar to the Korean case; the Korean performance management system also has a rule that top executives may be fired if they get a D or E grade over 2 consecutive years.

Validity of Performance Indicators

As noted earlier, the performance indicators that this paper utilizes are systemically evaluated; each indicator is evaluated quantitatively or qualitatively. Reports, which form the basis of the evaluation, should be audited by an external audit firm where independent committees grade it based on scores. However, there are several potential threats to the data. These potential threats are related to whether the data reflect the true performance of public organizations. To analyze this threat, this paper reviews different dimensions of public organizations' performances. These performances have two broad dimensions: perceptual and archival performance (Amirkhanyan et al. 2018).

Perceptual performance measures are based on the view of managers and employees. This type of performance informs us about its self-report scales and enables us to measure its various facets, something that archival performance cannot capture (Meier and O'Toole 2013a). However, perceptual performance has several problems. The first problem is positivity bias; employees and managers are more likely to view their performance favorably (Boyne et al. 2006). The second problem is the common method bias, which is a systematic bias that occurs when several similar variables are in one source (Meier and O'Toole 2013b). For instance, if various questions ask about performance in one survey, respondents may care about the consistency of their response rather than their real performance.

The performance evaluation of Korean government corporations may avoid shortcomings from the perceptual performance measure. This is because these government corporations are evaluated based on reports that they submit before their performance

evaluation; these reports include how performance is evaluated, such as using a formula. Furthermore, their performance is evaluated by independent committees. This means that employees themselves would have no room to measure their performance; therefore, the performance evaluation of Korean government corporations is relatively free from positivity bias and common method bias, which are both problems of the perceptual performance measure. Admittedly, a qualitative evaluation should be subjectively measured by independent committees. However, because independent committees are not employees of government corporations, the propensity for positivity bias and common method bias is relatively low. In addition, a final qualitative score is decided through consultation of committees. This may enhance the reliability of the performance evaluation.

Archival performance estimates achievement based on archival data, which are collected before the performance evaluation. This type of performance minimizes discretion and organizational bias (Amirkhanyan et al. 2018). However, archival performance may be problematic in that it may reflect the preferences of powerful actors by adjusting the compositions of performance measures and causing gaming behaviors, like cheating and manipulation.

The performance evaluation of Korean government corporations may also avoid problems from the influence of powerful actors. This evaluation is based on predetermined goals and formulas. This prevents the influence of powerful actors, since they may have difficulty predicting the environment in which their organizations are evaluated. Powerful actors may also change within the fiscal year. Additionally, government corporations set their performance targets by mutual consent with an independent committee; therefore, powerful actors are less likely to intervene in the process of formula setting. In summary,

a performance indicator that uses predetermined formulas formed by mutual consent, can minimize bias in the evaluation process and the intervention of powerful actors.

Empirical Strategy

This paper utilizes Regression Discontinuity Design to capture any significant changes in the level of employee fraud, at the cutoff points of organizational performance ratings. Since each government corporation has been assigned both scores and grades, the dataset for the performance evaluation of Korean government corporations gives a unique opportunity to researchers to estimate the impact of performance feedback. This setting, where each government corporation has a continuous score and a clear-cut final grade, allows this study to use a Regression Discontinuity (RD) design as the primary methodology to provide credible causal inferences about how performance affects subsequent fraudulent behaviors of employees. Intuitively, RD enables a comparison between government corporations where performance is just above the grade cutoff and where it is just below the grade cutoff. For instance, government corporations that received a standardized score of 0 are very similar to others that achieved a standardized score of -0.1. Minor differences in performance grades and scores may be caused by random factors, meaning these organizations are less likely to have meaningful differences. However, the performance signal between these two agencies is substantially different; the government corporation that achieved a standardized score of 0 gets a C grade, but the government corporation that received a -0.1 receives a D grade. Stakeholders, including staff, managers, regulators, and citizens, may think differently about an agency with a C grade, which has

an acceptable level of performance, compared to an agency with a D grade, which is not acceptable. Furthermore, RD can overcome possible serious bias that occurred in the OLS model. The legitimate concern in the OLS model is that high-performing organizations may have fundamental differences from low-performing organizations, but the OLS model does not effectively control these unobservable factors (Hong et al. 2020). This problem motivates the usage of a RD design, which assumes that the distribution of unobservable factors that cause biases in the OLS model changes smoothly across performance grades (Skovron and Titiunik 2015). Based on this assumption, the RD design examines whether a significant difference exists among organizations that are just above and just below the cutoff.

Data and Method

Operationalization of Employee fraud

To examine how low-performance signals impact employee fraud, this study leverages the results of integrity assessment for 104 South Korean public corporations over 3 years between 2015 and 2017. Measurement of employee fraud comes from subset scores of the internal integrity index[8]. It is especially based on survey items of budget operations. The Korean Anti-Corruption and Civil Right Commission (ACRC), annually evaluates the integrity of government corporations. Integrity assessment was introduced in 2002 to

[8] Although the final outcomes of the integrity assessment are publicly available at the website of ACRC (https://www.acrc.go.kr/), the results of each survey item are kept unpublished to the public and provided only for each government corporate. The 3-year data from 2015 to 2017 were provided by each government corporation with permission to use it strictly for research purposes.

government departments, local governments, and government corporations to estimate and compare the levels of organizational integrity to encourage public organizations to reduce corruption. The assessment includes two categories: external integrity and internal integrity. The evaluation of internal integrity measures organizational ethical climates including culture, personnel management, budget operation, and fairness in the work process. While the external integrity index is measured based on a citizen survey, the internal integrity index is estimated based on a public employee survey. This evaluation[9] is based on corruption events and a survey of citizens and public employees. Details of measurement are presented in Table 1. The items of budget operation consist of operating expenses, benefit allowances, and project expenses, and the survey provides examples of what behaviors violate each expense policy. Dependent variables in this study take the average of survey questions. For instance, the variable of abuse of operating expense is the average score of three items related to the concept of abusing operating expense. In front of each category in the survey, examples are given to improve understanding of what the abuse of operating expenses, benefit allowance, and project expenses are. The result of each item consists of a scale of 0 to 10 as same the way to output the final outcomes of the integrity assessment. The score is the average of employee responses, and it is aggregated to the organizational level. The final measure of employee fraud measures employees' perception of fraudulent behaviors, aggregated to the organization level.

[9] These results weight averages to produce a final result on a scale of 0-10. Typically, this survey usually uses a seven-point Likert scale; for example, questions asking unethical practices consist of every time, usually, frequently, sometimes, occasionally, rarely, and never. The responses are transformed to fit this scale, such as 1 for a score of 0.00 (or 10.00), 4 for a score of 5.00, and 7 for a score of 10.00 (or 0.00) so that a high score is an indicator that unethical behavior is well-managed in a certain government corporation. In 2015, approximately 170,000 citizens and 57,000 public employees participated in these surveys. In 2016, about 160,000 citizens and 55,000 public employees participated.

Table 1 Measurement of Dependent Variables

Dependent Variable	Item	
Abuse of operating expense	Q1	Are there any cases of unjustifiably executing operating expenses in your organization for a year?
	Q2	How many cases are there unjustifiably spent operating expenses for a year?
	Q3	How much expense is there unjustifiably spent on operating expenses for a year?
Abuse of a benefit allowance	Q1	Are there any cases of unjustifiably requesting extra pay and/or travel funds in your organization for a year?
	Q2	How many cases are there unjustifiably requested extra pay and/or travel funds for a year?
	Q3	How much expense is there unjustifiably requested extra pay and/or travel funds for a year?
Abuse of project expense	Q1	Are there any cases of unjustifiably executing project funds in your organization for a year?
	Q2	How many cases are there unjustifiably executed project funds for a year?
	Q3	How much expense is there unjustifiably executed project funds for a year?

Table 2 presents the examples of the concepts. These examples are provided in front of the survey item to improve the understanding of survey participants. First, examples of violating operating expense policy cover the behaviors where employees resell gift cards to make cash and treat someone else who is not related to the task. Second, cases of violating benefit allowance policy include the behaviors where employees inflate their working hours and budget for business trips. Third, abuse of project expense in the survey

indicates private use of project expenses by inflating expected costs and revenue assignment to whom it has not participated in the project. In addition, it is also essential to assess the reliability and validity of measurements to index multiple questionnaires (Wang & Kim, 2022). This study, therefore, conducts post hoc assessments including Cronbach's alpha and factor analysis to check internal consistency and dimensionality. Tables 3 and 4 show the factor loadings and Cronbach's alpha.

Table 2 Example to improve understanding of survey questions

Dependent Variable	Example
Abuse of operating expense	• Buying some private products with operating expense • Using operating expenses for a private gathering • Buying gift cards and reselling for encashment • Using operating expenses for external stakeholders or acquaintances to support their activities
Abuse of a benefit allowance	• Requesting travel funds by inflating the days of travel • Requesting overtime pay even if they did not work at that time
Abuse of project expense	• Requesting salary by registering employees who did not actually participate in the project • Private use of project funds by inflating the price of items • Using the balance of project funds for other purposes

Table 3 Factor Loadings of Items over the Study Period

		Factor Loadings

	Item	2014	2015	2016
Abuse of operating expense	Q1	0.90	0.92	0.93
	Q2	0.92	0.94	0.96
	Q3	0.94	0.90	0.91
Abuse of a benefit allowance	Q1	0.91	0.86	0.93
	Q2	0.92	0.94	0.96
	Q3	0.88	0.92	0.93
Abuse of project expense	Q1	0.91	0.83	0.88
	Q2	0.92	0.90	0.93
	Q3	0.84	0.85	0.83

Table 4 Cronbach's Alpha over the Study Period

	Cronbach's Alpha		
Item	2014	2015	2016
Abuse of operating expense	0.91	0.91	0.93
Abuse of a benefit allowance	0.89	0.89	0.93
Abuse of project expense	0.87	0.82	0.85

RD Approach

The RD design in this study uses a global polynomial regression that compares just below and just above the cutoff with the continuity of the average. That is, the global RD model uses all observations to estimate the impact of performance signals on the degree of employee fraud. The reason that this study utilizes the global model is because the small number of observations. The local RD model requires sufficient observations around the cutoff point for rigorous estimation; estimation would be biased when the local RD model

is applied with a small size of observation. In this sense, this study, which relies on a relatively small size of observation for empirical analysis, employs a global RD model to minimize the bias of the RD estimator. The global RD model in this study is as follows:

$$Y_{s(t+1)} = a + f\left(X_{s(t)}\right) + \beta G_{s(t)} + \gamma Covariates_t + \pi_t + \epsilon_{st}$$

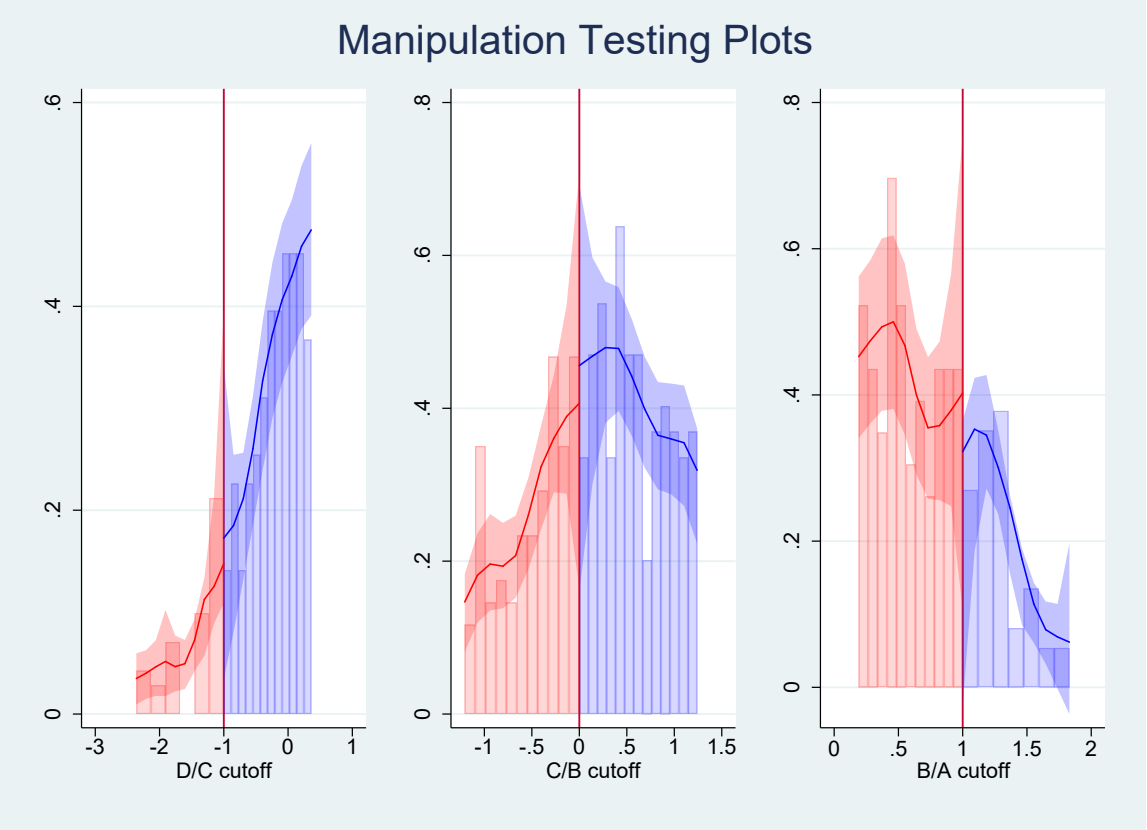
Where $Y_{s(t+1)}$ indicates the indicators of employee fraud in government corporation s in year t+1. $f(X_{st})$ is a smooth function of the running variable, including linear, quadratic, and cubic polynomial forms in year t. $G_{s(t)}$ is the grade determined by the running variable; this study sets a C/D grade as a primary cutoff point, which is a junction that divides average and poor performances (Hong 2019). The vector of covariates consists of the characteristics of human resources and financial resources. Variables of human resources include average wage, entry-level wage, the number of employees, and the number of job openings. Characteristics of financial resources cover the size of the asset, dept, expenditure, administration cost, labor cost, and government subsidy. π_t are the time-fixed effects.

Validation of the RD design

This study conducts validation tests to validate the RD model. To utilize RD design appropriately, the model should hold three assumptions: no manipulation of grades, unique cutoff, and homogenous characteristics except for treatment status (Cattaneo et al., 2019).

First, units should not have the capacity to alter the value of the running variable to receive certain performance ratings. Regression discontinuity design assumes arbitrary cutoffs and checks the treatment effect around cutoffs. The capacity that observations purposely manipulate the value of running variables, hence, weaken the validity of RD design. Second, there should only be discontinuities at the true cutoff points. If significant discontinuities occur at the other points, it could threaten the validity of the treatment effect in RD results. Third, covariates should be balanced between treated and control units. It means that significant discontinuity should be presented in dependent variables and all covariates should continue smoothly over the cutoffs. Techniques of checking assumptions follow the ways of Cattaneo et al., (2019).

To check the first assumption, this paper conducts the local polynomial density test, which checks the significant discontinuity of observation density at the cutoff points. This test assumes that there are significant jumps and drops at the cutoff points if observations can manipulate their performance scores. Figure 1 and Table 4 indicate that observations are not likely to manipulate their values in the running variables. The 95% confidence intervals overlap at the primary cutoffs in Figure 1 and all p-values are bigger than 0.05 in Table 4.



Note: Line and shading area indicate the point estimate and 95% confidence interval respectively

Figure 1 Density Plots at Cutoffs

Table 4 Manipulation Test Results

Cutoff	(1) D / C	(2) C / B	(3) B / A
T-value	-0.610	-0.055	-1.303
P-value	0.54	0.96	0.19
Effective N	48	103	65

Note: The null hypothesis is that there is no difference in the density of treated and control observations at the cutoff

This paper employs placebo tests to check the second assumption of whether the primary cutoff is unique. Placebo tests run the RD regression with false cutoffs and check the smooth relationship between the control and treatment groups. This study sets 4

alternative cutoff points below and above the original C/D grade cutoff, which is a primary cutoff in the RD model. With the suggestions of Cattaneo et al., (2019), the tests include observations only in the control group to check the discontinuity of the points below the original cutoff and observations only in the treatment group for the test with points above the original cutoff. Table 5 shows the results of placebo tests with each outcome variable. Although there are significant effects in terms of abuse of operating expense and benefits allowance at the point of -0.7, they are only significant at the 10% confidence level and smaller than the conventional threshold for statistical significance.

Table 5 Results with placebo cutoffs near Original C/D cutoff

	(1)	(2)	(3)
	Abuse of operating expense	**Abuse of a benefit allowance**	**Abuse of project expense**
Alternative cutoff : **-1.30**	-3.935 [-3.94 1.75]	-2.918 [-6.66 3.2]	-2.25 [-6.21 2.4]
N	40	40	40
Alternative cutoff : **-1.15**	-3.972 [-20.12 7.26]	-.433 [-20.62 30.69]	-2.857 [-3.19 22.05]
N	40	40	40
Alternative cutoff : **-0.85**	-1.273 [-4.02 3.62]	-1.19 [-4.09 3.35]	-1.683 [-3.57 1.8]
N	272	272	272
Alternative cutoff : **-0.70**	-2.095* [-3.4 .27]	-1.482* [-3.85 .23]	-.933 [-3 1.23]
N	272	272	272

To check the third assumption, this study conducts balance tests with covariates. The reason being that fundamental concept of the RD model is to leverage as-if random

assignments at the cutoff points. Treated and control units should be very similar with the exception of the treatment status. This study, therefore, tests the smooth continuity at the cutoff based on covariates in the model. Table 6 shows the RD results based on the model setting the control variables as outcome variables. At the primary cutoff, which is the C/D cutoff in this study, there are no significant discontinuities at any polynomial forms. It indicates that the treatment and control groups share similar characteristics at the baseline.

Table 6 Balance Check with Covariates at the Cutoff between D and C Grade

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Num. of employees	Num. of job opening	Asset	Dept	Government subsidy	Labor cost	G&A cost	Expenditure	Wage	Entry level wage
			(Billion Won)	(Billion Won)	(Billion Won)	(Billion Won)	(Billion Won)	(Billion Won)	(Million Won)	(Million Won)
RD estimator	-642.24	-28.99	-3555	-1578	-2411	-52	-51	-3305	-6.5	-1.3
(Linear form)	[-1618.4 293.4]	[-79.9 11.5]	[-17076 8526]	[-13046 8733]	[-7268 3223]	[-115 25]	[-145 51]	[-9259 3920]	[-18 3.4]	[-5.6 2.8]
N	312	312	312	312	312	312	312	312	312	312
RD estimator	-662.48	-34.23	-4275	-2157	-2014	-45	-47	-2670	-7.2	-1.4
(Quadratic form)	[-1959.4 440]	[-96.6 18.4]	[-23919 13903]	[-10976 5329]	[-1.24, 0.44]	[-143 34]	[-188 35]	[-13613 6373]	[-23 2.3]	[-7.2 3.6]
N	312	312	312	312	312	312	312	312	312	312
RD estimator	-759.71	-39.12	-5008	-1580	-2808	-5457	-77	-3620	-10	-1.8
(Cubic form)	[-1597.8 1230.4]	[-104.6 49.1]	[-18707 24894]	[-14670 23861]	[-11400 8174]	[-122 84]	[-234 6]	[-13950 9772]	[-24 3.8]	[-9.3 2.8]
N	312	312	312	312	312	312	312	312	312	312

Note: Inference based on robust bias corrected confidence intervals that are presented in brackets. All models include year fixed effects and covariates. Standard errors clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Results

Descriptive Statistics

Table 7 presents summary statistics of dependent and control variables. All dependent variables show stable results over the study period. As noted earlier, a scale of 0 to 10 indicates the degree of employee fraud. A score of 0 means all subjects of the survey perceived that there were employee fraud in their organizations while a score of 10 means all of them did not agree. Dependent variables, presented in Table 7, were balanced over the study period. Abuse of operating expense and benefits allowance scored approximately 7, which indicates employees perceive that those fraudulent behaviors of employees occasionally occur in their organizations. The score of violating project expense policy was above 8, which means that there rarely occurred in public corporates in the research period. Most covariates show increasing patterns over the years; the business size, costs, wages, and even liabilities of government corporates gradually increased compared to the baseline. For example, the annual average income in 2014 was 65,000,000 Korean won, which is equivalent to 55,000 USD, and it grew to 67,000,000 Korean won in 2016. As another example, each government corporate hired approximately 73 employees in 2014 and 85 employees in 2016.

Table 7 Descriptive Statistics

		2014 (N = 104)	2015 (N = 104)	2016 (N = 104)
Dependent variable				
Abuse of operating expense	Mean	7.52 (2.45)	6.59 (2.75)	6.83 (2.71)

Abuse of a benefit allowance	Mean	7.60 (2.28)	6.85 (2.60)	7.02 (2.70)
Abuse of project expense	Mean	8.60 (2.41)	8.19 (2.38)	8.14 (2.61)
Control variable				
Num. of employees	Mean	1,402.9 (2,615.64)	1,431.65 (2,660.23)	1,505.62 (2,758.05)
Num. of job openings	Mean	72.65 (118.45)	67.60 (125.86)	84.71 (170.62)
Asset (Billion Korean Won)	Mean	7,399 (24,700)	7,780 (26,100)	8,089 (26,900)
Dept (Billion Korean Won)	Mean	4,949 (18,400)	5,171 (19,100)	5,284 (19,500)
Government subsidy (Billion Korean Won)	Mean	3,099 (9,358)	3,530 (11,100)	3,225 (9,835)
Labor cost (Billion Korean Won)	Mean	92 (181)	100 (198)	107 (213)
G&A cost (Billion Korean Won)	Mean	68 (144)	70 (143)	77 (173)
Expenditure (Billion Korean Won)	Mean	3,982 (10,800)	4,464 (12,400)	4,101 (11,200)
Wage (Million Korean Won)	Mean	65 (12)	67 (13)	69 (14)
Entry-level wage (Million Korean Won)	Mean	32 (4.5)	33 (4.6)	34 (4.6)

Figure 2 presents the distribution of performance ratings in each year. As noted earlier, performance ratings for Korean government corporates were assigned by standard deviations converted from raw scores (Hong, 2019b). Over half of the government corporates lie between B and C grades. Grade D, which is the primary cutoff in this study, is given only if the score is below -1 standard deviation while grade A is assigned if the score is above +1 standard deviation. The percentages of each grade from E/D to A were

approximately 13%, 31%, 41%, and 14% in 2014, 13%, 22%, 46%, and 19% in 2015, and 13%, 32%, 41%, and 14% in 2016.

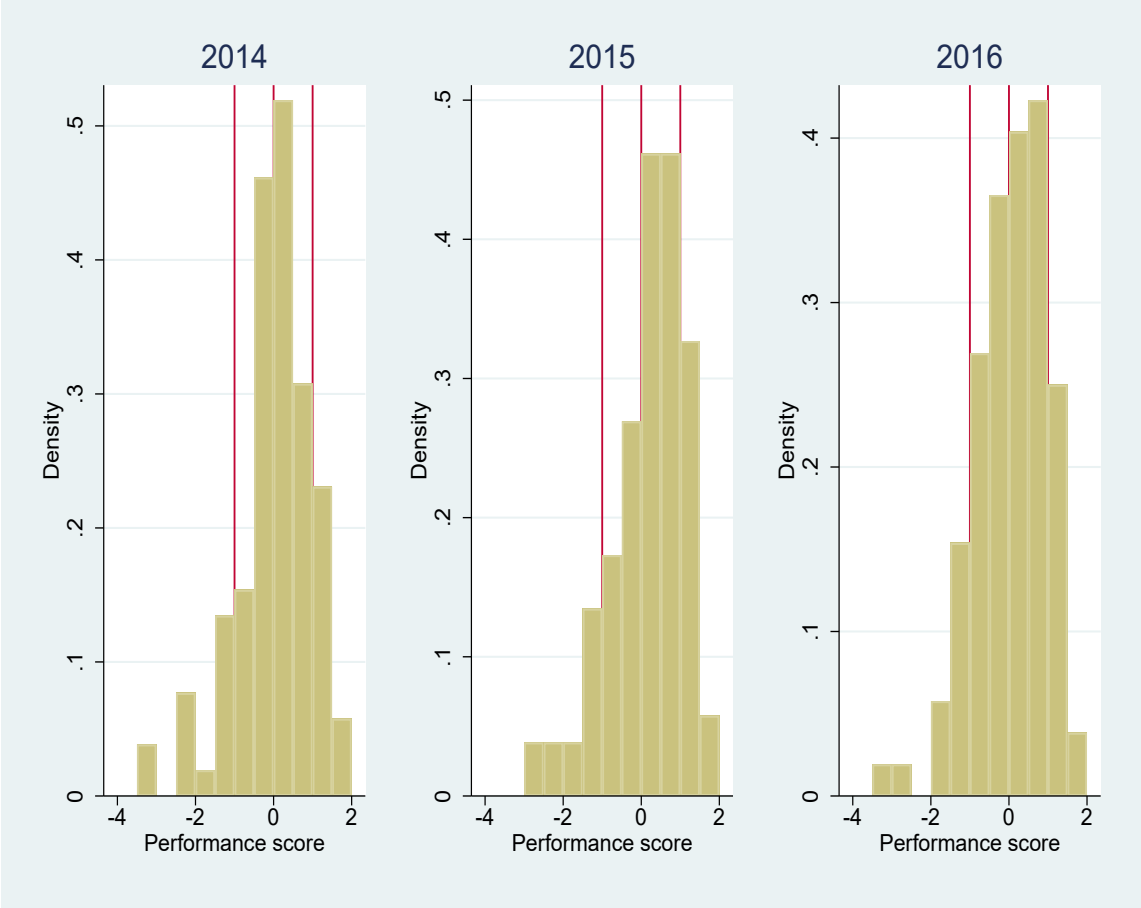
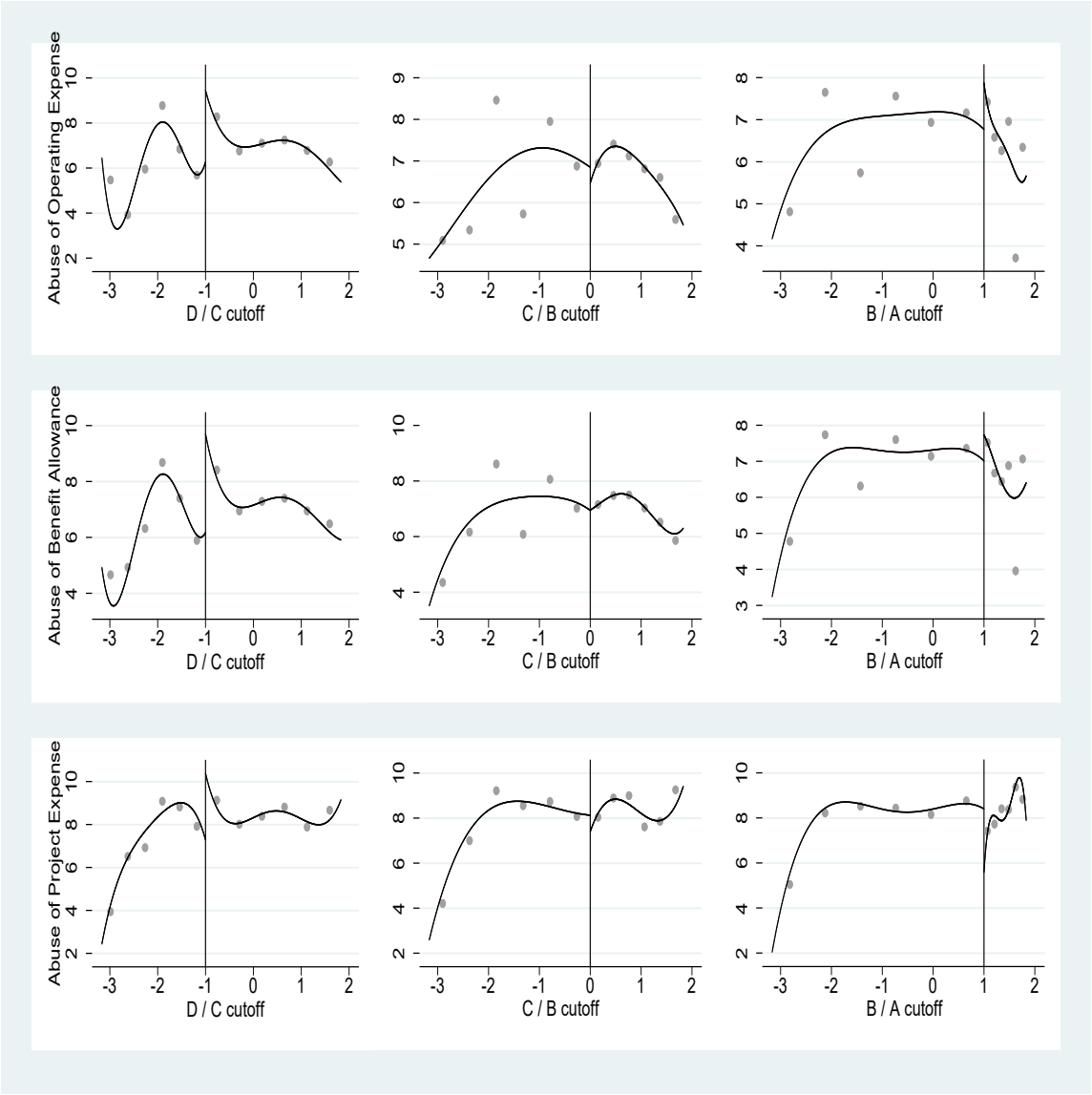


Figure 2 Distribution of Performance Scores and Grades

As graphical inspection is an essential part of the RD model, this study presents the RD plot to check the discrete jumps or drops of outcome variables at each cutoff. Figure 3 shows there are clear and large discontinuities at the C/D cutoff in all dependent variables. The scores to measure employee fraud jumped at the C/D cutoff; it means government corporates with D and E grades show lower scores of employee fraud. On the other hand,

there are relatively small gaps at the B/C and A/B cutoff; it lets us infer the discontinuity at the C/D cutoff is highly robust.



Note: The dots indicate local averages of scores for each dependent variable. Polynomial fit order of 4

Figure 1 RD plots for Employee fraud

The RD estimates in Table 8 support the results of the RD plot which we visually analyzed with a 4th-order polynomial trend. The scores measuring employee fraud

significantly jumped at the C/D cutoff while there was no significant change at the other cutoffs. It suggests that employees are more likely to perceive more frequent employee fraud in response to poor-performance signals than average- and high-performance signals. In terms of operating expense and benefits allowance, by the forms of polynomials, government corporates with grades above D increased scores from 2 to 4 at the 95% or 99% confidence level. The results are consistent across alternative polynomial functional forms. In contrast, there was no statistical significance at the cutoff between B/C and between A/B. The results based on project expense are similar to the results of the previous two dependent variables. When organizations received poor performance ratings, employees in those organizations are more likely to recognize the violation of project expense policy. Although the results of project expense were significant at the A/B cutoff, the estimations with various polynomial forms were not consistent. The result with linear form was not significant while the result with quadratic form was only significant at the 90% confidence level. Overall, with those estimations, employee fraud related to fictitious and inflating expenses may become prevalent in response to low-performance signals.

Table 8 RD Regression Results for Employee Fraud

Dependent variables	(1) Linear	(2) Quadratic	(3) Cubic
Abuse of operating expense			
RD estimator	1.939***	3.122***	4.428**
(D/C cutoff)	[1.62 4.6]	[2.49 6.27]	[1.2 5.72]
N	312	312	312
RD estimator	.005	-.437	-.477
(C/B cutoff)	[-1.83 1.13]	[-2.49 1.54]	[-3.11 2.02]
N	312	312	312
RD estimator	.52	1.086	.759
(B/A cutoff)	[-.54 2.65]	[-1.2 2.73]	[-1.6 2.84]
N	312	312	312
Abuse of a benefit allowance			
RD estimator	1.741***	3.107***	4.414***
(D/C cutoff)	[1.59 4.59]	[2.37 6.37]	[1.35 6.21]
N	312	312	312
RD estimator	.088	-.243	-.443
(C/B cutoff)	[-1.58 1.25]	[-2.33 1.45]	[-2.6 2.32]
N	312	312	312
RD estimator	.334	1.04	.302
(B/A cutoff)	[-.59 2.59]	[-1.7 2.34]	[-2.11 2.4]
N	312	312	312
Abuse of project expense			
RD estimator	.557***	1.805***	2.961***
(D/C cutoff)	[.35 3.22]	[1.18 4.66]	[1.27 5.51]
N	312	312	312
RD estimator	-.071	-.055	-1.069
(C/B cutoff)	[-1.36 1.35]	[-2.74 .65]	[-3.15 1.18]
N	312	312	312
RD estimator	-1.196	-1.566*	-2.386**
(B/A cutoff)	[-3.65 .59]	[-4.86 .12]	[-5.57 -.85]
N	312	312	312

Discussion and Conclusion

To summarize, this study examines the causal relationship of how low-performance signals raise the risk of employee fraud with three indicators. First, the degree of operating expense policy violation was higher at low-performing organizations in the subsequent year.

Employees in low-performing organizations are more likely to observe abuse of budget for general activities, such as purchasing equipment and gift cards, to make cash for themselves. Second, low-performing organizations are more likely to experience cases of violation of the benefit allowance policy. More employees answered that their organizations have cases inflating the costs of business trips and working hours. Third, the scores related to project expense were low at poor-performing organizations as well. Employees in low-performing organizations are more likely to learn the cases of unjustified spending on project management including favoritism and artificial inflation on project costs.

These findings require scholars and practitioners to take seriously the relationship between low performance and employee fraud. First, this study confirms the necessity of monitoring the behaviors of low-performing organizations. In addition to evidence of the organizational cheating in performance management (Bohte & Meier, 2000b; Greve, 2003; Hong, 2019b; Wang, 2022), this study provides evidence that low-performance signals are related to employee fraud. The evidence of this study allows us to infer that low-performance signals may trigger not only organizational cheating but also employee fraud. Therefore, it may remind regulators of the necessity of monitoring the behaviors in low-performing organizations. On the other hand, the findings also recall the discussion of the accountability system and pressures from the system. Because the findings come from performance management designed with a reward-and-sanction accountability system (Hong, 2019b), we could expand our discussion to the accountability system and pressure. Accountability system and pressure provide reasons for the manager to rush to improve their performance; the expected sanction on managerial authorities amplifies the negativity

bias, which leads managers to react sensitively to performance shortfalls (Hong & Kim, 2019; Kogan et al., 2016; van den Bekerom et al., 2021). Ratcheting up accountability pressure, which is expected to positively affect only performance outcomes, paradoxically, increases the risk of unethical conducts in public organizations (Han & Wang, 2022). Overall, the argument of this study links the logic of a vicious cycle that flows from performance shortfalls to occurrences of unethical conducts to poor performance. Recalling the previous literature on ethic management, gaming behaviors, and performance management, ratcheting up performance pressure increases the risk of unethical behaviors (Heckman et al., 2002; Heinrich & Marschke, 2010; Moynihan, 2009), and unethical climates negatively related to performance outcomes (Hassan et al., 2014; Khaltar & Moon, 2020).

Findings allow us to infer the potential solutions for employee fraud as well. A balance between accountability pressure and managerial authorities could be a potential solution to reduce employee fraud in public organizations. As noted earlier, the conventional rewards-and-sanctions system imposes restrictions on managerial authority, such as human and financial resources, as a punishment for poor performance (Hong, 2019b; Wang & Yeung, 2019). Limited managerial authorities may threaten the leadership to control the organizations and lower the number of options that managers can utilize in the process of improving performance outcomes (Fairholm, 2004; Han & Wang, 2022). Considering that ethical leadership is key to preventing employees' unethical behaviors in a workplace (DeMatthews & Serafini, 2019; Downe et al., 2016; Hassan, 2019), sufficient room to demonstrate managers' leadership could reduce employee fraud. In addition, it raises doubts about the effectiveness of restricting financial incentives on employees in

poor-performing organizations. Different from the private sector, public employees are less likely to be incentivized by pay-for-performance (Gillam et al., 2012; Minchin et al., 2018). If pay-for-performance does not work for the service quality and the gap in financial rewards would increase the risk of employee fraud (Bajorek & Bevan, 2015; Peters & Maniam, 2016), it may be necessary to reconsider the role of financial rewards in the public sector.

Despite thoroughly analyzing the relationship between performance signals and employee fraud with RD design, there are some limitations. A notable one is that the measurements of dependent variables rely on the perceptions of employees. Even though each index includes the survey items asking the number of occurrences and the size of financial values, the answers are based on the subjects' memories. It could cause distortion to grasp the relationship appropriately. However, it does not mean that it is not worth using perpetual data. Because unveiled unethical behaviors are only a small number in the practice, scholars often suffer from the shortage of archival data to do research on unethical behaviors (Belle & Cantarelli, 2017). In this sense, the perpetual measure of employee fraud could complement the research with archival data. Another limitation is generalizability. This limitation comes from the fact that this study depends on the case of government corporates and performance management with a rewards-and-sanctions system. While government corporates become familiar with public management in tandem of New Public Management (Eger III, 2006; Frant, 1997), the several characteristics of government corporates, such as earning profits and requiring pay in return for service, differentiate from conventional organizations in the government sector. Moreover, because the discussion of this study depends on the concepts of rewards-and-sanctions

accountability, it may not be answerable how low-performance signals affect employee fraud in other accountability systems.

This study, nevertheless, makes several contributions to both theory and practice. First, it directly contributes to the literature on employee fraud and performance management; employee fraud is a novel type of unethical behavior related to performance management. It deepens our understanding of how performance shortfalls and associated organizational cheating and lack of financial reward affect employees' fraudulent behavior. Additionally, this study sheds light on the different patterns of employee fraud by the level of performance. The RD model uses a set of fine-grained cutoffs by performance levels and provides nuances that the patterns of employee fraud are differentiated from the performance signals. Practically, this study helps regulators and managers to refine the performance management system to lower the occurrence of fraudulent behaviors in public organizations. The empirical evidence that low-performance signals cause subsequent employee fraud informs the expected effect of low-performance signals and helps to take action before employee fraud occurs.

Reference

Adler, P. S., & Borys, B. (1996). Two types of bureaucracy: Enabling and coercive. *Administrative Science Quarterly*, 61–89.

Bajorek, Z. M., & Bevan, S. M. (2015). Performance-related-pay in the UK public sector: A review of the recent evidence on effectiveness and value for money. *Journal of Organizational Effectiveness: People and Performance*.

Belle, N., & Cantarelli, P. (2017). What Causes Unethical Behavior? A Meta-Analysis to Set an Agenda for Public Administration Research. *Public Administration Review*, *77*(3), 327–339. https://doi.org/10.1111/puar.12714

Benaine, S. L., & Kroll, A. (2020). Explaining effort substitution in performance systems: The role of task demands and mission orientation. *Public Management Review*, *22*(6), 813–835. https://doi.org/10.1080/14719037.2019.1604794

Berman, E. M., & West, J. P. (1997). Managing ethics to improve performance and build trust. *Public Integrity Annual*, *2*(1), 21–29.

Bevan, G., & Hood, C. (2006). WHAT'S MEASURED IS WHAT MATTERS: TARGETS AND GAMING IN THE ENGLISH PUBLIC HEALTH CARE SYSTEM. *Public Administration*, *84*(3), 517–538. https://doi.org/10.1111/j.1467-9299.2006.00600.x

Bhatti, W. A., Waris, S., Zaheer, A., & others. (2011). THE EFFECT OF COMMITMENT AND MOTIVATION ON HUMAN TALENT AND ITS CONTRIBUTION TO ORGANIZATIONAL PERFORMANCE. *Management & Marketing*, *6*(3).

Blazovich, J. L., Smith, K. T., & Smith, M. (2013). Employee-friendly companies and work-life balance: Is there an impact on financial performance and risk level? *Journal of Organizational Culture, Communications and Conflict, Forthcoming*.

Bohte, J., & Meier, K. J. (2000). Goal Displacement: Assessing the Motivation for Organizational Cheating. *Public Administration Review*, *60*(2), 173–182. https://doi.org/10.1111/0033-3352.00075

Brignall, S., & Modell, S. (2000). An institutional perspective on performance measurement and management in the 'new public sector.' *Management Accounting Research*, *11*(3), 281–306. https://doi.org/10.1006/mare.2000.0136

Burke, F., & Black, A. (1990). Improving organizational productivity: Add ethics. *Public Productivity & Management Review*, 121–133.

Case, J. (2000). *Employee theft: The profit killer*. John Case & Associates.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). A Practical Introduction to Regression Discontinuity Designs: Foundations. *Elements in Quantitative and Computational Methods for the Social Sciences*. https://doi.org/10.1017/9781108684606

Chen, S. (2023). Performance feedback, blame avoidance, and data manipulation in the public sector: Evidence from China's official city air quality ranking. *Public Management Review*, 1–21.

Chun, Y. H. (2005). Goal Ambiguity and Organizational Performance in U.S. Federal Agencies. *Journal of Public Administration Research and Theory*, *15*(4), 529–557. https://doi.org/10.1093/jopart/mui030

DeMatthews, D. E., & Serafini, A. (2019). Do Good Principals Do Bad Things? Examining Bounds Of Ethical Behavior In The Context Of High-Stakes Accountability. *Leadership and Policy in Schools*, 1–20. https://doi.org/10.1080/15700763.2019.1668023

Dizon-Ross, R. (2020). How does school accountability affect teachers? Evidence from New York City. *Journal of Human Resources*, *55*(1), 76–118.

Downe, J., Cowell, R., & Morgan, K. (2016). What Determines Ethical Behavior in Public Organizations: Is It Rules or Leadership? *Public Administration Review*, *76*(6), 898–909. https://doi.org/10.1111/puar.12562

Eger III, R. J. (2006). Casting light on shadow government: A typological approach. *Journal of Public Administration Research and Theory*, *16*(1), 125–137.

Fairholm, M. R. (2004). Different Perspectives on the Practice of Leadership. *Public Administration Review*, *64*(5), 577–590. https://doi.org/10.1111/j.1540-6210.2004.00405.x

Fox, S., Spector, P. E., & Miles, D. (2001). Counterproductive work behavior (CWB) in response to job stressors and organizational justice: Some mediator and moderator tests for autonomy and emotions. *Journal of Vocational Behavior*, *59*(3), 291–309.

Frant, H. (1997). Reconsidering the determinants of public authority use. *Journal of Public Administration Research and Theory*, *7*(4), 571–590.

Gillam, S. J., Siriwardena, A. N., & Steel, N. (2012). Pay-for-Performance in the United Kingdom: Impact of the Quality and Outcomes Framework--A Systematic Review. *The Annals of Family Medicine*, *10*(5), 461–468. https://doi.org/10.1370/afm.1377

Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel. *Psychological Science*, *20*(3), 393–398. https://doi.org/10.1111/j.1467-9280.2009.02306.x

Goddard, M., Mannion, R., & Smith, P. (2000). Enhancing performance in health care: A theoretical perspective on agency and the role of information. *Health Economics*, *9*(2), 95–107. https://doi.org/10.1002/(SICI)1099-1050(200003)9:2<95::AID-HEC488>3.0.CO;2-A

Goh, E., & Kong, S. (2018). Theft in the hotel workplace: Exploring frontline employees' perceptions towards hotel employee theft. *Tourism and Hospitality Research*, *18*(4), 442–455. https://doi.org/10.1177/1467358416683770

Greenberg, J. (1990). Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology*, *75*(5), 561.

Greenberg, J. (1997). *A social influence model of employee theft: Beyond the fraud triangle.*

Greenberg, J. (2002). Who stole the money, and when? Individual and situational determinants of employee theft. *Organizational Behavior and Human Decision Processes*, *89*(1), 985–1003. https://doi.org/10.1016/S0749-5978(02)00039-0

Greve, H. (2003). *Organizational Learning from Performance Feedback*. Cambridge.

Han, X., & Wang, W. (2022). Does granting managerial autonomy in exchange for accountability mitigate gaming? *Public Administration Review*, puar.13564. https://doi.org/10.1111/puar.13564

Hassan, S. (2019). We Need More Research on Unethical Leadership Behavior in Public Organizations. *Public Integrity*, *21*(6), 553–556. https://doi.org/10.1080/10999922.2019.1667666

Hassan, S., Wright, B. E., & Yukl, G. (2014). Does Ethical Leadership Matter in Government? Effects on Organizational Commitment, Absenteeism, and Willingness to Report Ethical Problems. *Public Administration Review*, *74*(3), 333–343. https://doi.org/10.1111/puar.12216

Heckman, J. J., Heinrich, C. J., & Smith, J. A. (2002). *The performance of performance standards*. National Bureau of Economic Research Cambridge, Mass., USA.

Heinrich, C. J., & Marschke, G. (2010). Incentives and their dynamics in public sector performance management systems. *Journal of Policy Analysis and Management*, *29*(1), 183–208. https://doi.org/10.1002/pam.20484

Holm, J. M. (2018). Successful Problem Solvers? Managerial Performance Information Use to Improve Low Organizational Performance. *Journal of Public Administration Research and Theory*, *28*(3), 303–320. https://doi.org/10.1093/jopart/muy017

Hong, S. (2019). A Behavioral Model of Public Organizations: Bounded Rationality, Performance Feedback, and Negativity Bias. *Journal of Public Administration Research and Theory*, *29*(1), 1–17. https://doi.org/10.1093/jopart/muy048

Hong, S., & Kim, Y. (2019). Loyalty or Competence: Political Use of Performance Information and Negativity Bias. *Public Administration Review*, *79*(6), 829–840. https://doi.org/10.1111/puar.13108

Hood, C. (2007). What happens when transparency meets blame-avoidance? *Public Management Review*, *9*(2), 191–210. https://doi.org/10.1080/14719030701340275

Jacobsen, C. B., & Bøgh Andersen, L. (2017). Leading public service organizations: How to obtain high employee self-efficacy and organizational performance. *Public Management Review*, *19*(2), 253–273. https://doi.org/10.1080/14719037.2016.1153705

Johnson, W. C., Xie, W., & Yi, S. (2014). Corporate fraud and the value of reputations in the product market. *Journal of Corporate Finance*.

Julnes, P. de L., & Holzer, M. (2001). Promoting the Utilization of Performance Measures in Public Organizations: An Empirical Study of Factors Affecting Adoption and Implementation. *Public Administration Review*, *61*(6), 693–708. https://doi.org/10.1111/0033-3352.00140

Kelman, S., & Friedman, J. N. (2009). Performance Improvement and Performance Dysfunction: An Empirical Examination of Distortionary Impacts of the Emergency Room Wait-Time Target in the English National Health Service. *Journal of Public*

Administration Research and Theory, *19*(4), 917–946. https://doi.org/10.1093/jopart/mun028

Khaltar, O., & Moon, M. J. (2020). Effects of Ethics and Performance Management on Organizational Performance in the Public Sector. *Public Integrity*, *22*(4), 372–394. https://doi.org/10.1080/10999922.2019.1615163

Kim, S. (2004). Individual-Level Factors and Organizational Performance in Government Organizations. *Journal of Public Administration Research and Theory*, *15*(2), 245–261. https://doi.org/10.1093/jopart/mui013

Kogan, V., Lavertu, S., & Peskowitz, Z. (2016). Performance Federalism and Local Democracy: Theory and Evidence from School Tax Referenda: PERFORMANCE FEDERALISM. *American Journal of Political Science*, *60*(2), 418–435. https://doi.org/10.1111/ajps.12184

Koning, P., & Heinrich, C. J. (2013). Cream-skimming, parking and other intended and unintended effects of high-powered, performance-based contracts. *Journal of Policy Analysis and Management*, *32*(3), 461–483.

Kroll, A., & Vogel, D. (2021). Why public employees manipulate performance data: Prosocial impact, job stress, and red tape. *International Public Management Journal*, *24*(2), 164–182. https://doi.org/10.1080/10967494.2021.1874579

Kuenzi, M., Brown, M. E., Mayer, D. M., & Priesemuth, M. (2019). Supervisor-Subordinate (Dis)agreement on Ethical Leadership: An Investigation of its Antecedents and Relationship to Organizational Deviance. *Business Ethics Quarterly*, *29*(1), 25–53. https://doi.org/10.1017/beq.2018.14

Lynn, L. E. (1996). *Public management as art, science, and profession*. Chatham House Publishers Chatham, NJ.

Marcus, B., & Schuler, H. (2004). Antecedents of counterproductive behavior at work: A general perspective. *Journal of Applied Psychology*, *89*(4), 647.

Meier, K. J., Favero, N., & Zhu, L. (2015). Performance Gaps and Managerial Decisions: A Bayesian Decision Theory of Managerial Action. *Journal of Public Administration Research and Theory*, *25*(4), 1221–1246. https://doi.org/10.1093/jopart/muu054

Menzel, D. C. (2015). The ethics factor in local government: An empirical analysis. In *Ethics and Public Administation* (pp. 191–204). Routledge.

Menzel, D. C. (2016). State of the art of empirical research on ethics and integrity in governance. *Ethics in Public Management*, 24–54.

Min, B. H., Oh, Y., & Brower, R. S. (2021). The Effects of Diverse Feedback Dynamics on Performance Improvement: A Typology of Performance Feedback Signals. *Administration & Society*, *53*(1), 123–157. https://doi.org/10.1177/0095399720933822

Minchin, M., Roland, M., Richardson, J., Rowark, S., & Guthrie, B. (2018). Quality of Care in the United Kingdom after Removal of Financial Incentives. *New England Journal of Medicine*, *379*(10), 948–957. https://doi.org/10.1056/NEJMsa1801495

Moynihan, D. P. (2009). Through A Glass, Darkly: Understanding the Effects of Performance Regimes. *Public Performance & Management Review*, *32*(4), 592–603. https://doi.org/10.2753/PMR1530-9576320409

Nicholson-Crotty et al. (2017). *Performance and Management in the Public Sector: Testing a Model of Relative Risk Aversion:* 12.

Nielsen, P. A. (2014). LEARNING FROM PERFORMANCE FEEDBACK: PERFORMANCE INFORMATION, ASPIRATION LEVELS, AND MANAGERIAL PRIORITIES: Learning from performance feedback. *Public Administration*, *92*(1), 142–160. https://doi.org/10.1111/padm.12050

Owusu, G. M. Y., Koomson, T. A. A., Alipoe, S. A., & Kani, Y. A. (2022). Examining the predictors of fraud in state-owned enterprises: An application of the fraud triangle theory. *Journal of Money Laundering Control*, *25*(2), 427–444. https://doi.org/10.1108/JMLC-05-2021-0053

Peters, S., & Maniam, B. (2016). Corporate fraud and employee theft: Impacts and costs on business. *Journal of Business and Behavioral Sciences*, *28*(2), 104.

Portz, J. (2021). "Next-Generation" Accountability? Evidence From Three School Districts. *Urban Education*, *56*(8), 1297–1327. https://doi.org/10.1177/0042085917741727

Rickman, N., & Witt, R. (2007). The Determinants of Employee Crime in the UK. *Economica*, *74*(293), 161–175. https://doi.org/10.1111/j.1468-0335.2006.00530.x

Rustiarini, N. W., T., S., Nurkholis, N., & Andayani, W. (2019). Why people commit public procurement fraud? The fraud diamond view. *Journal of Public Procurement*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/JOPP-02-2019-0012

Salge, T. O. (2011). A Behavioral Model of Innovative Search: Evidence from Public Hospital Services. *Journal of Public Administration Research and Theory*, *21*(1), 181–210. https://doi.org/10.1093/jopart/muq017

Sohn, H., & Bae, K. B. (2018). Do Public Employees "Game" Performance Budgeting Systems? Evidence From the Program Assessment Rating Tool in Korea. *The American Review of Public Administration*, *48*(5), 458–475. https://doi.org/10.1177/0275074016689322

Taylor, J. (2009). STRENGTHENING THE LINK BETWEEN PERFORMANCE MEASUREMENT AND DECISION MAKING. *Public Administration*, *87*(4), 853–871. https://doi.org/10.1111/j.1467-9299.2009.01788.x

Taylor, J. (2021). Public Officials' Gaming of Performance Measures and Targets: The Nexus between Motivation and Opportunity. *Public Performance & Management Review*, *44*(2), 272–293. https://doi.org/10.1080/15309576.2020.1744454

Thaler, J., & Helmig, B. (2016). Do Codes of Conduct and Ethical Leadership Influence Public Employees' Attitudes and Behaviours? An experimental analysis. *Public Management Review*, *18*(9), 1365–1399. https://doi.org/10.1080/14719037.2015.1103890

van den Bekerom, P., van der Voet, J., & Christensen, J. (2021). Are Citizens More Negative About Failing Service Delivery by Public Than Private Organizations? Evidence From a Large-Scale Survey Experiment. *Journal of Public Administration Research and Theory*, *31*(1), 128–149. https://doi.org/10.1093/jopart/muaa027

Van der Voet, J. (2022). Policy preferences in response to negative economic prospects of COVID-19: A survey-experiment among local politicians in four European countries. *Public Administration*, *100*(2), 356–374.

van der Voet, J. (2023). Search in Response to Negative Performance Feedback: Problem-Definition and Solution-Generation. *Journal of Public Administration Research and Theory*, *33*(1), 43–62. https://doi.org/10.1093/jopart/muac008

Wang, W. (2022). How does performance management affect social equity? Evidence from New York City public schools. *Public Administration Review*, puar.13590. https://doi.org/10.1111/puar.13590

Wang, W., & Kim, T. K. (2022). Examining the effects of a performance management reform on employee attitudes and organizational climate. *Public Management Review*, 1–23. https://doi.org/10.1080/14719037.2022.2026095

Wang, W., & Yeung, R. (2019). Testing the Effectiveness of "Managing for Results": Evidence from an Education Policy Innovation in New York City. *Journal of Public*

Administration Research and Theory, *29*(1), 84–100. https://doi.org/10.1093/jopart/muy043

Zhu, L., & Rutherford, A. (2019). Managing the Gaps: How Performance Gaps Shape Managerial Decision Making. *Public Performance & Management Review*, *42*(5), 1029–1061. https://doi.org/10.1080/15309576.2019.1568886

VITA

Taek Kyu Kim was born in Seoul, Korea, on October 26, 1987. He attended Sewon High School in Goyang, Korea, and graduated in February 2006. After he found the additional opportunity to take the university entrance exam, he entered Kyung Hee University, Seoul, Korea in March 2007. In 2014, he and his colleagues received Best Undergraduate Student Paper Award from Korean Association for Public Administration for their research on the generational conflicts in public transportation. In the following February, he received the degree of Bachelor of Arts in public administration. By working as a research assistant of faculties and researchers at Kyung Hee University in 2016, he became assured that he was interested in and talented in researching public administration and public policy topics. He entered Yonsei University, Korea in March 2016 and received a Master of Arts in public administration in February 2018. During this period, he achieved several honorable awards including Excellence Local Finance Paper Award in 2016, Rhee Yang Su Fellowship Award in Spring and Fall 2017, Certificate of Distinction in 2017, and Best Graduate Student Paper Award in 2017. To experience a diverse environment and improve his capacity, he decided to further his studies in the United States. He entered the Truman School of Public Affairs, University of Missouri, as a recipient of a G. Ellsworth Huggins Fellowship which is a 5-year doctoral-level fellowship. In 2022, he was honored Staats Emerging Scholar Award by the Networks of Schools of Public Policy, Affairs, and Administration (NASPAA). He is expected to receive a Doctor of Philosophy in public affairs in July 2023. Taek kyu will join the University of Exeter in the United Kingdom as an Assistant Professor in Public Policy and Public Administration in September 2023.