ABSTRACT

Title of Dissertation: THE IMPACT of PERFORMANCE MANAGEMENT ON DEMOCRATIC VALUES: EVIDENCE FROM US EDUCATION

Xu Han Doctor of Philosophy, 2022

Dissertation directed by: Dr. Toby M. Egan, Associate Professor

Although performance management is supposed to be a generic, values-neutral tool that can be adapted for any purpose, it has been criticized for ignoring important democratic values. Critics claim that violating these democratic values makes performance management counterproductive to the stated aim of restoring public trust in government through improved outcomes.

This dissertation comprehensively examines the impact of performance management on equity and civic engagement in U.S public high schools. Performance management may incentivize prioritizing high-value students who are more likely to contribute to school performance ranking at the expense of others, creating an inequity problem. However, it can also promote the well-being of disadvantaged groups by providing incentives and information on improvement in disaggregated performance. Performance management may draw resources and attention to activities aiming to improve students' academic performance in high-stakes subjects (reading, math, and science) at the expense of other important activities where students develop skills in and interests for civic engagement. However, activities aiming to improve students' academic performance also prepare students to perform tasks such as reading, writing, speaking, and quantitative reasoning, integral parts of civic engagement.

To conduct the analysis, the dissertation draws on a nationally representative survey of administrators and students at public high schools. As students' academic performance is the result of collaborative efforts among students and staff (teachers and principals), performance management is operationalized for students and staff respectively. The student component includes

established student performance standards, frequency of standardized testing, and imposed consequences. The staff component includes principals' managerial autonomy, teachers' evaluation, and imposed consequences.

Through a multilevel analysis of how performance management influences students, especially for racial minorities' standardized test scores in math, findings point to an unfilled promise regarding equity. Performance management components for students and staff are each associated with increased average student test scores, but do not shrink the gap between advantaged and disadvantaged student subgroups. Still, not all aspects of performance management perpetuate inequity. Of the two performance management components focusing on staff and students, the staff component is associated with lower test scores for struggling students while the student component increases struggling student performance.

By simultaneously analyzing the indirect effects of performance management on volunteering behaviors through cognitive abilities, civic skills, and civic norms in structural equation modeling, the dissertation finds mixed effects of performance management on civic engagement. On the one hand, the student component has a positive but small indirect effect on civic participation by improving students' cognitive abilities. On the other hand, the staff component has a negative but small indirect effect by reducing students' participation in extracurricular activities where they develop civic skills. However, the student component does not negatively affect civic engagement. Overall, the findings suggest that despite the negative effects of performance management on equity and civic engagement, performance management can be used to mitigate inequity and reverse the recent decline in civic engagement.

THE IMPACT OF PERFORMANCE MANAGEMENT ON DEMOCRATIC VALUES: EVIDENCE FROM US EDUCATION

by

Xu Han

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Professor Toby M. Egan, Chair
Professor Jeffery Harring
Professor Donald Moynihan
Professor Philip Joyce
Professor Peter Reuter

TABLE OF CONTENTS

LIST OF TABLES

TABLE OF FIGURES

INTRODUCTION

Although performance management is supposed to be a generic tool that can be adapted for any purpose and, therefore, value-neutral, it has been criticized for ignoring important democratic values and having become counterproductive to restoring public trust in government through improvement in performance (Radin 2006; Snyder, Saultz, and Jacobsen 2017). Democratic values refer to the values conducive to striking a balance between freedom and equity, the ultimate goal of a just and democratic society (Rawls 2001). At the center of this criticism is the possible trade-off between measured and unmeasured results. While performance management can be effective in improving measured results, the improvements are often realized at the expense of unmeasured but equally important results. Public education serves as an ideal place to study the tensions not least because multiple and possibly competing values and goals coexist in public education (R. Rothstein, Jacobsen, and Wilder 2008). Like any other public organization, citizens expect public schools to be efficiently run, but they also care about disadvantaged groups and the negative consequences efficiency-based decisions can bring to these vulnerable groups. In addition, citizens expect schools to not only equip students with economically valuable skills to compete in the global economy but also civic skills to participate in the governance of communities.

While these goals can coexist with each other, performance management is claimed to narrow attention toward some public education goals at the expense of others (Snyder, Saultz, and Jacobsen 2017). For instance, high-stakes testing is blamed for crowding out teachers' internal motivation, leaving teachers to be driven primarily by external motivations such as performance bonuses, promotions, and sanctions. As performance measurement does not usually gauge all valuable activities, using performance management as the primary motivator

will drive people to spend more effort on what is measured at the expense of unmeasured but equally valued activities (Dixit 2002; Holmstrom and Milgrom 1991). Effort substitution behaviors include focusing on students close to the performance targets, who are referred to as bubble kids (Booher-Jennings 2005), at the expense of others and teaching to the test at the expense of students' all-around development (Hirsh-Pasek and Steinberg 2018; Rothstein, Jacobsen, and Wilder 2008). These claimed negative effects of performance management feed the belief that what gets measured are often gamed or manipulated.

These gaming behaviors depart from democratic values. Gaming behaviors are behaviors that improve measured behaviors without advancing the objectives behind the performance metrics (Benaine and Kroll 2020). For example, the practice of focusing on "bubble kids", children whose performance is close to the proficiency level (Booher-Jennings 2005), contradicts the value of equity and its requirement to allocate resources to students who need the help most. Ignoring individual student development in the pursuit of increasing collective test scores can hinder the development of these future citizens' abilities to participate in and deliberate over public affairs. The resulting skills deficit serves as a potential barrier to students from fulfilling their long-term rights and civic duties. The possible contradictions between current approaches to performance management and democratic values warrant consideration. These deviations will undermine public trust and defeat the claimed purpose of performance management, to restore public trust in government (Wichowsky and Moynihan 2008). Further encroachment on trust in government and citizens' unfulfilled civic duties are detrimental to the health of future democracy.

There is a dearth of large-N studies on these important topics. To address the gap in large N studies, this dissertation examines the effect of performance management on equity and

democratic citizenship, two pillars of democratic values. Equity is critical to realizing the ideal that everyone has equal moral worth (Rawls 2001). Citizens are expected to participate in public affairs in not only an equal but also productive manner. To achieve this goal, citizens need to be active in and skilled at civic participation. Specifically, Study 1 examines the influences of performance management on equity by studying the effects of performance management on student subgroup math scores. Study 2 gauges the influence of performance management on democratic citizenship by testing the effects of performance management on students' civic participation.

Performance management refers to a system where performance information created through strategic planning, and routine measurement is connected with and ideally influences decisions (Moynihan 2008). In education, a performance management system has been extensively applied to improve student academic performance (Mehta 2013). The emphasis on the effectiveness and efficiency of improving students' academic performance in performance management often clashes with other important educational goals such as fostering citizenship and advancing equity. Hence, public education provides an ideal setting for examining how performance management shapes democratic values.

Pathways from Performance Management to Performance Improvement

There are three pathways for performance management to deliver performance improvement. Performance management provides managers with the information, incentive, and autonomy to deliver performance improvement. Each of the pathways is reviewed in the following section so that readers can better appreciate the importance of comprehensively measuring performance management, an approach taken in this dissertation.

Information Pathway

Performance management informs the allocation of attention and resources. By specifying concrete and challenging goals, performance management can motivate public employees to strive for better performance (Locke and Latham 2002; Favero, Meier, and O'Toole 2016). In the implementation phase, performance management provides feedback that helps managers identify problems and search for solutions (Behn 2014; Holm 2018). At the end of performance cycle, disclosing comparative performance information (how organizations perform in comparison with their peers and the past) will also prompt public managers to improve performance (Hong 2019; Nielsen 2014a). Last but not least, evidence of program ineffectiveness may make managers to revisit assumptions underlying existing programs and revise strategic planning. Using performance information for continually updating strategic planning is associated with higher organizational performance, especially at the mature stage of performance management (Han and Moynihan 2021). In public education, performance management has also been associated with better management of teachers (Rockoff et al. 2012) and improvement in school academic achievement (Dee and Wyckoff 2015).

Incentive Pathway

Performance management can incentivize performance improvement by attaching consequences to performance goals. Expectancy theory holds that individuals can be motivated towards goals if (1) expectancy that efforts lead to desired outcomes is strengthened, (2) expectation of a reward for meeting performance goals is enhanced (instrumentality), and (3) the values individuals place on rewards is increased (valence) (Vroom 1964). By enhancing the perceived linkage between performance and reward, performance management can motivate public employees to strive for accomplishing goals.

Some scholars question the effectiveness of incentive pathway in improving organizational performance for at least three reasons. First, public organizational performance is subject to more external influences (O'Toole and Meier 2011). So, it is challenging to attribute performance changes to actions or inaction of a specific unit in public organizations. Second, the link between performance and reward can be tenuous in systems claiming to use performance as the basis for rewards (Heinrich and Choi 2007). Third, intrinsic motivation such as public service motivation is the primary motivation for most public employees in the United States (Perry, Engbers, and Jun 2009). So, attaching consequences to performance goals can be either ineffective or even counterproductive for motivating public employees (Frey, Homberg, and Osterloh 2013). In public education, the record of motivating teachers and administrators by connecting their performance with rewards has been mixed (Fryer 2017). Much appears to depend on the specific incentive design and setting in which they are imposed (Goodman and Turner 2013; Imberman and Lovenheim 2015).

Autonomy Pathway

Performance management empowers managers to manage for results. To respond to the incentives and utilize performance feedback, managers need to be granted autonomy in managing organizations (Moynihan 2008; Ouchi and Segal 2003). Empowering managers enables managers to adjust programs according to performance feedback and quickly address the performance deficit (Behn 2014; Holm 2018). The empowerment also makes it possible for public managers to tailor programs to citizens' needs, improving the chance of successfully serving citizens (Ouchi and Segal 2003). In public education, granting managers autonomy in exchange for accountability to results has also been shown to improve school performance in various settings (D. Osborne 2017; Wang and Yeung 2018).

Gaming In Performance Management

Gaming is an umbrella term that includes all types of perversity in performance management, ranging from data manipulation, effort reduction to effort substitution (Benaine and Kroll 2020; Heinrich and Marschke 2010; Kelman and Friedman 2009). How performance management affects effort substitution is a key area of focus in this dissertation. The section covers how performance management may generate incentives and opportunities for gaming (Taylor 2021). The review of gaming behaviors in performance management will demonstrate why the dissertation incorporates performance management for various stakeholders in the provision of public services.

Poor designs of performance management systems generate incentives for gaming the system in at least three ways. First, attaching consequences to narrowly defined performance metrics will make public employees overlook other high-stakes but equally valuable goals, such as equity and civic participation (Radin 2006; Wichowsky and Moynihan 2008). It is also worth noting that the criticism of performance management's encroachment on civic participation is often based on anecdotes and theoretical discussion. Few studies have empirically verified this claim.

Second, rigid goal setting and unrealistic targets often leaves public employees with little choice but to game the system (Terman and Yang 2016). One example of unrealistic goal setting is the failure to take into account the influence of other stakeholders on program outcomes. Hence, the characteristics and behaviors of citizens served by public programs will influence the difficulty of achieving the same level of performance improvement (D. Osborne 2017; O'Toole and Meier 2011). The task difficulty will, in turn, influence the incentive for organizations to prioritize improvement in measured performance at the expense of others (Benaine and Kroll

2020). Consequently, a performance management system that fails to tailor to service users' needs and characteristics may miss their performance target and increase the pressure for gaming the system. In other contexts, unrealistic goal setting also leads to data manipulation to demonstrate goal attainment (Gao 2020).

Third, public managers are often held accountable to results without being granted autonomy and resources for delivering results. Although the doctrine of performance management holds that managers need to have autonomy to manage for results in exchange for being held accountable for results (Moynihan 2008), public managers are often held accountable to results without obtaining the necessary authority to achieve positive changes (Jakobsen and Mortensen 2016). Even if managers receive the autonomy for managing the daily operation of organizations, they receive few resources or support to deal with increasing complexity and volumes of tasks. As a result, inexperienced managers may perform worse after the empowerment (Han and Wang 2021). To demonstrate goal attainment and to cope with relentless performance pressure in this circumstance public employees may engage in either effort substitution, effort reduction, or data manipulation (Tummers et al. 2015; Soss, Fording, and Schram 2011).

Performance management systems may also provide public employees more opportunities to game the system. Performance management marks a significant shift from process-oriented accountability to results-oriented accountability. The shift often generates more performance measures in a wide range of areas (Moynihan 2008). While availability of performance measures makes it possible to measure results of government programs, it also exacerbates the principal-agent challenge of verifying and interpreting data. The explosion of data provides cover for public employees' selective manipulation of data (Jacob and Levitt

2003). The availability of diverse performance metrics also makes it easier to engage in motivated reasoning: select and present favorable metrics in support of one's preexisting opinions (James et al. 2020). Last but not least, it is often not clear what a performance indicator implies (Joyce 2007). For instance, increasing, decreasing funding can be equally valid conclusions for a program with poor performance. The ambiguity in interpreting performance information and the common requirement of evidence-based policymaking creates more opportunities for spinning performance information.

Gaps in the Literature

Among extant research focusing on the effects of performance management on democratic values in public education, key gaps exist. The literature has (1) inconsistent and narrow operationalizations of performance management. For the question of the effects of performance management on democratic citizenship, there is simply (2) a lack of empirical large N studies to comprehensively test the effects of performance management on democratic citizenship.

First, the effect of performance management on organizational performance has been extensively studied, but the operationalization of performance management is inconsistent and often too narrow to completely capture the full tenets of performance management. Policy analyses of performance management often focus on the incentive pathway (Dee and Jacob 2011; Krieg 2008; Neal and Schanzenbach 2010; Reback 2008; Springer 2008). Public management research largely studies the information pathway (Andersen and Nielsen 2019; Boyne and Chen 2008; Gerrish and Spreen 2017; Poister, Pasha, and Edwards 2013). With a few exceptions, the autonomy pathway is often left out of studies in quantitative analysis of performance management.

The lack of consistent and comprehensive operationalization of performance management poses challenges to the study of performance management. It is important to holistically study performance management systems. Research on strategic human resource management (HRM) holds that as HRM practices can have interaction effects with each other (complementary or contradictory), considering a system of HRM practices can often yield a more systemic understanding of the impacts of HRM on performance. This emphasis on a bundle of practices also applies to performance management.

Specifically, a narrow measurement of performance management can impede progress in understanding the impact of performance management in at least two ways. First, components of performance management are often complementary to each other. Continuous performance feedback is needed to guide the direction of improvement for public employees held accountable for performance targets; managers need to be empowered to manage for results in exchange for being held accountable for results (Moynihan 2008; Ouchi and Segal 2003). Studying one piece of performance management components, such as attaching consequences to performance-related information, may lead to difficulty in understanding the null effect of the specific component. If all the other complementary components are in place, then the null effect can be interpreted as evidence that performance management does not work. If some important complementary parts of performance management are missing, then the null effect indicates a poor implementation of performance management.

Second, a focus on performance management for either service users or service providers may lead to an incomplete understanding of what drives performance in public service. The results of public service delivery often depend on collaboration between service users and service providers (Osborne, Radnor, and Strokosch 2016). In public education, the

success of improving student educational outcomes relies on not only what public employees do but also on how students do at schools (Ouchi and Segal 2003). Examining performance management for only service providers or service users can also lead to an incomplete picture of how performance management shapes stakeholders' behaviors and organizational performance (Springer et al. 2010; 2012). Hence, it is important to measure performance tracking and application of performance consequences for both parties to better diagnose reasons for existing performance.

Third, negative effects of performance on democratic citizenship have been well-documented in qualitative studies (Hirsh-Pasek and Steinberg 2018; R. Rothstein, Jacobsen, and Wilder 2008), the negative effects from these small N studies may not tell the whole story. Focusing on reading and math test-based accountability as extensions of performance management strengthens students' abilities to participate in public affairs. Moreover, resource models for civic participation predict that people with strong civic skills and high cognitive abilities will find participation easier and will tend to participate at higher levels (Musick and Wilson 2008; Verba, Schlozman, and Brady 1995). By improving students' reading and math scores, both valid measures of cognitive abilities (Dee and Jacob 2011; Hanushek and Woessmann 2010; Favero, Meier, and O'Toole 2016), performance management can enhance civic participation through positive effects on cognitive abilities. For this reason, a study can be unbalanced if it only examines the alleged negative effects of performance management on student civic participation.

Methodological Overview

Principal component analysis (PCA) was used to capture the variation in the measures of performance management ranging from goal setting, and performance disclosure, to attaching consequences to performance (Mulaik 2010). PCA showed that two components of performance

management were sufficient to capture the majority of variation. This reduction (from six to two measures) significantly reduces the complexity of moderation analysis in Study 1.

The multilevel model was utilized to study how performance management moderates the negative association of racial minority, limited English proficiency, and special education status with students' math scores. I made this choice for three reasons. First, the multilevel model is also suitable for studying how the variation in school management (macro-level unit) influences the students' performance (micro-level unit) (Snijders and Bosker 2012). Second, the multilevel model controlled for the correlation among students at schools and provides an accurate estimate for the standard error. Third, controlling for both individual and school aggregated measures of demographic variables avoids ecology fallacy and leads to accurate estimates of the relationship between demographic variables and student math performance at the individual level.

To enhance the internal validity of findings from study one, two types of robustness checks were conducted. First, to control for the selection bias that leaders of performance management may be different from laggards of performance management, inverse probability weighting was used to give higher weights to schools which are similar in other aspects except in the level of implementing performance management (Mccaffrey et al. 2013). Second, to control for the between-state variation in implementing accountability systems, measures for the between-state variation in disaggregated performance reporting, attaching consequences to performance (Education Week 2003), and prior experience for implementing test-based accountability (Hanushek and Raymond 2005) were incorporated into models.

Structural equation models (SEMs) were used to estimate the indirect relationships between performance management and civic participation for two reasons. First, SEMs simultaneously estimated the associations of performance management with civic participation

through multiple pathways ranging from cognitive abilities, and civic skills, to the endorsement of citizenship norms (Byrne 2011). Second, SEMs controlled for the correlations between the pathways. Specifically, the development of civic skills was proxied by both the width and depth of participation in extracurricular activities. SEMs controled for the correlation between these two pathways in the models.

To enhance the internal validity of findings in study 2, four robustness checks were conducted. First, to mitigate the concern for reverse causality between participation in extracurricular activities and volunteering, proxy measures of civic participation, student volunteering behaviors in the base period (2002) was included and the dependent variable of volunteering measures in 2004 was replaced with the same ones in 2006. The second robustness check used state dummies to control for the between state variation in the requirement of community service (Shapiro 2018) and verified whether the effects of predictors for students' volunteering behaviors still held. The third test studied whether the results remained robust after taking into account people's tendency for responding to survey questions regardless of question content. The fourth test distinguished between within-school variation and between school variation and estimated the indirect effect of school-centered measures and school averaged measures of math scores, frequency, types of extracurricular activities, and endorsement of civic norms on volunteering behaviors (Zhang, Zyphur, and Preacher 2009).

Contributions

This dissertation aims to advance understanding regarding the effects of performance management on organizational performance in three ways by—(1) measuring goal setting, performance monitoring, and attaching consequences to both service users and providers; (2) aspiring to be the first large-N empirical study about the effects of performance management

on democratic citizenship; (3) advancing policy feedback theory (Moynihan and Soss 2014) by providing the first empirical evidence regarding how managerial reforms resulting from policy changes shape citizens' capabilities and preferences for participation; and (4) exploring the conditions under which performance management can mitigate gaming behaviors and strengthen democratic values.

First, the dissertation directly measures the variation in goal setting, performance monitoring, and attaching consequences to performance for both service users and service providers. In addition, the operationalization of performance management includes a component of managerial empowerment, a key tenet of performance management (Moynihan 2008; Wang and Yeung 2018). This operationalization captures almost all the tenets of performance management and the coproduction nature of public education. Study 1 examines how the variations in the components of school performance management moderated the effect of individual characteristics that are traditionally associated with lower math scores such as previous low performance, racial minority, low socioeconomic status, limited English proficiency, and special education status.

Second, structural equation modeling (SEM) was used to study the effects of performance management on civic participation in Study 2, constituting the first attempt to empirically test the effect of performance management on citizens' preference for and capability in participation. SEM is a "statistical technique that takes a confirmatory approach to the analysis of structural theories bearing on some phenomenon" (Byrne 2012, 3). Compared with traditional multivariate regression, SEM can directly calculate and control for the measurement errors, deal with complex causal relationships such as indirect effects and include unobserved variables (i.e. latent variables indicated by several measured variables) (Byrne 2012; Kline 2015). SEM was

used to simultaneously analyze the indirect effects of performance management on writing scores through reading scores, civic participation through math knowledge and skills, civic skills and individual endorsement of citizenship norms. It also controlled for the other commonly cited predictors for civic participation—such as relative socioeconomic status, racial heterogeneity, and social capital (Nie, Junn, and Stehlik-Barry 1996; Putnam 2000, 2007).

Third, Study 2 further develops policy feedback theory by testing the postulation that policy affects politics through its effect on public management. Policy feedback theory holds that policy is not just an output of politics, but it, in turn, shapes politics (Mettler and Soss 2004). The dissertation tests a new channel through which policy influences politics. Literature on policy feedback theory often focuses on how the details of specific programs such as means-testing vs. universal eligibility influence people's participatory behaviors and attitudes (Campbell 2012; Soss 1999, 2002; Kumlin and Rothstein 2005). However, policy is not self-executing. Bureaucracy implements policy and is influenced by the process of implementation (Skocpol 1992). It is possible that the effect of policy on public management will have a second-order effect on people's political behaviors and attitudes (Moynihan and Soss 2014). Policies emphasizing the economic importance of school performance and accountability to school performance have brought performance management to the center of school management (Gross et al. 2005; Massell et al. 2005; Mehta 2013). Still, schools vary in implementing performance management. This study aims to examine how this variation in performance management affects students' (citizens) civic participation by shaping their preference and capacity for participation.

Fourth, the dissertation draws on bundle theory in strategic human resource management to explore conditions under which performance management can mitigate gaming and enhance democratic values. The current literature on gaming in performance management focuses on how

to improve the policy design to mitigate gaming (Neal 2011). The managerial solutions to gaming are less understood. The policy solutions alone are likely to be insufficient because policy designs may not anticipate all the possible scenarios and design fixes in advance (Holmstrom and Milgrom 1991). To address this gap, I applied bundle theory to explore how management can mitigate gaming behaviors and strengthen democratic values. The theory of HRM bundles holds that HRM practices will be satisfying to employees and improve organizational performance if practices are internally coherent (horizontal fit), aligned with organizational strategy (vertical fit), and compatible with the organizational system including resources, structure, and culture (organizational fit) and external environment (environment fit).

In the discussion section, the ways in which the framework can help public managers improve performance management and mitigate gaming are demonstrated. The framework will guide public managers to evaluate (1) whether the information, incentive, and autonomy pathways for performance improvement are impeded in the implementation of performance management (horizontal fit), (2) whether practices of performance management are aligned with the stage of performance management (vertical fit), (3) whether resources are sufficient to ensure full implementation of performance management (organizational fit), and (4) whether the external environment is compatible with the requirements of performance management (environmental fit). Together, this will contribute to a better understanding of how performance management can mitigate gaming and strengthen democratic values.

PERFORMANCE MANAGEMENT IN PUBLIC EDUCATION

Since the publication of *A Nation at Risk* in 1983, the US has witnessed the gradual shift to a new paradigm emphasizing the economic significance of education, which includes—the importance of comprehensive improvement; the responsibility of schools, rather than society, to

realize the improvement; and the use of performance measures through test scores (Mehta 2013). This paradigm brings performance management to the center of school management and culminated in the enactment of No Child Left Behind Act (NCLB), a nationwide test-based accountability system, in 2001.

Improving American School Act (IASA) of 1994 is the first national act aimed at moving K-12 public schools toward performance management. IASA required states to establish challenging content and performance standards, track students' performance against these standards, and hold schools and school systems accountable for students' performance. To this end, it mandated standardized tests in math and reading at least once during each of three grade spans: 3rd-to-5th, 6th-to-9th, and 10th-to-12th (Goertz 2005). Since then, many states have held schools accountable for not only aggregated outcomes such as school test scores, but also disaggregated outcomes like the school test scores by racial groups (Hanushek and Raymond 2005).

The No Child Left Behind Act (NCLB), which was passed in 2001 and implemented in 2002, advanced accountability based on standards and tests in three ways. First, NCLB increased the frequency of tests and required tests on reading and math annually from 3rd thru 8th grades. Meanwhile, it maintained the same frequency of tests at high schools as IASA. In addition, NCLB required states to test science annually in grades 3-5, 6-8, and 10-12. Science was tested, but NCLB mainly used the percentage of students reaching proficiency in math and reading to evaluate school progress. Schools meeting the required targets would be characterized as achieving Adequate Yearly Progress (AYP). Although AYP was introduced into federal law by IASA in 1994, it became a national high-stakes goal after the enactment of NCLB (Education Week 2011).

Second, NCLB expanded the categories of disaggregated performance goals to include low income, limited English proficiency, and special education status and required states to hold public schools accountable for these disaggregated performance goals. NCLB required each state to report AYP for both the student population and for particular "subgroups" of students, including English-learners and students in special education, racial minorities, and children from low-income families. NCLB mandated that every student reached proficiency level in reading and math by the school year 2013-2014.

Third, NCLB required states to mete out rewards and sanctions to all the public schools according to their performance in reaching or exceeding AYP standards. NCLB specified a series of escalating consequences for schools receiving Title I funds that fail to meet AYP (Education Week 2011). If a Title I school failed to make AYP for two consecutive years, the school needed to notify students' parents of the failure and offer the option of transferring to better-performing public schools in the same district. If the school failed to attain AYP for three years, it needed to provide supplementary services such as free tutoring. Missing AYP for four straight years resulted in state intervention such as school restructuring and even closure (Klein 2015). Despite many complaints about performance-based accountability, most schools took ongoing action to meet accountability requirements. These measures included efforts such as—additional after-school programs for test preparation, block schedules for test-related activities, and adjustment of curriculum and instruction practices (Gross et al. 2005; Massell et al. 2005).

To enable all students to learn from quality teachers and achieve higher performance, NCLB previously required that existing teachers must have a bachelor's degree and a license and/or certificate in the area they teach; new teachers must have bachelor's degree and pass the subject-matter exam (ASCD 2015). NCLB also stipulated that schools receiving Title 1-A

funding should prevent poor and minority students being taught by inexperienced, unqualified, or out-of-field teachers at a higher rate than other children. NCLB mandated that school provided professional development activities to teachers of core academic subjects.

Every Student Succeeds Act (ESSA), which President Obama signed into law in 2015, made several revisions to NCLB. First, it required states to adopt challenging standards in math, reading, and science (ASCD 2015). Second, it not only expanded academic measures to include school graduation rates but also included nonacademic factors chosen by states to measure school performance. Third, it eliminated AYP and the requirement of full proficiency within a certain time frame. Fourth, it forbade the federal government from prescribing specific school improvement strategies and grants states and districts the power to develop evidence-based strategies for improvement. Fifth, it removed the requirement for teachers to hold a bachelor's degree in the area they teach. Sixth, it expanded the professional development activities to teachers of all subject areas and other school staff.

Emphasis on goal setting, disaggregated data collection, collection, and accountability for results in NCLB continue in ESSA. First, ESSA recognized the importance of setting goals and standards, it moved one step further to aim for challenging goals and standards (U.S. Department of Education 2017). Second, ESSA maintained the reporting requirement of key academic performance metrics such as graduation rates and student subgroup test scores on the annual state report card (ASCD 2015). Third, diverse quality measures such as results from National Assessment on Education Progress (NAEP), postgraduation enrollment rates, and teachers' qualifications are also required to be disclosed on the annual report card. Fourth, districts and states are demanded to take actions to maintain accountability and help underperforming schools. Although AYP was eliminated, states remained committed to holding

schools accountable for results. They used diverse indicators, including academic performance and graduation rates, to identify a group of schools in need of support and improvement. Schools identified as in need of support are disclosed on the states' annual report cards. Once schools fall into the category of needing support, districts are expected to work with parents and schools to develop evidence-based strategies to improve outcomes. If the underperforming schools fail to meet the improvement criteria, the state is expected to enhance interventions in these schools. In conclusion, the current federal policy still maintains the same focus on collecting and disclosing performance information and holding schools accountable for performance. School management under NCLB epitomizes these emphases. Hence, studying performance management under NCLB will inform our understanding of current education policy.

The similarities and differences between NCLB and ESSA made it suitable to use Educational Longitudinal Study dataset from 2002 to 2006, the period under NCLB, to examine the relationship between equity and performance management. As the names of both legislations indicate, promoting equity and improving the academic performance of traditionally disadvantaged groups are the priorities in both. In addition, the major performance management components of goal setting, performance monitoring, and attaching consequences to performance all continue, albeit in different forms. Given the similar focus of policies and components of performance management, the success and failure of performance management in promoting equity can inform our current understanding of the relationship.

Moreover, the similarities and differences between NCLB and ESSA Educational longitudinal data enable me to examine the most likely case for the claim that performance management may develop the economic workforce at the expense of democratic citizenship.

NCLB marks the culmination of measuring performance and attaching consequences to performance in high-stakes academic subjects (Mehta 2013). Despite sharing an emphasis on goal setting, performance monitoring, and accountability to performance, ESSA eliminates the usage of adequate yearly progress as an accountability goal, requires the reporting of diverse quality measures, and reduced federal government involvement in applying consequences to performance (ASCD 2015). The limited and flexible approach toward school accountability and the reporting of more diverse quality measures make effort substitution from unmeasured aspect performance to measured one less likely. So, if performance management did not have a detrimental effect on civic participation at the height of school accountability, it is even less likely to have such a negative impact after ESSA was enacted

Studying public education is important for advancing knowledge of public administration. Public education ranks as the largest sector in terms of state and local governments' (SLGs) expenditures and the number of people employed by SLGs (Lewis and Pathak 2014; The Urban Institute-Brookings Institute Tax Policy Center 2017). Given the significance of public education in SLGs, public administration research focusing on public education is critical for ensuring a complete understanding of how government works in the US (Calista 2002; Raffel 2007). Moreover, employees' characteristics in public education such as high education attainment mirror those of public sector employees (Lewis and Pathak 2014; Mayer 2014; O'Toole, Jr. and Meier 2011). Findings from studies of the highly educated employees in public education is likely to be transferable to the general US public workforce because of their similarity in characteristics. Furthermore, the interconnectedness between public education and public administration suggests mutual benefits of integrating the study of public administration with that of public education. Public education is integral to many public

policy and management projects ranging from enhancing upward mobility and workforce diversity to economic development (Raffel 2007). As governors and mayors regained control of local education systems, more public administrators took on the top leadership positions in schools. The changing leadership composition at schools calls on students of public administration to expand their knowledge of public education. Enhancing public educators' knowledge of public administration can also help them buffer shocks such as environmental disasters, gain information, resources, and autonomy from the complex and fragmentary environment (Meier and O'Toole 2008; Meier, O'Toole, and Hicklin 2010; O'Toole and Meier 2011).

STUDY 1: CAN PERFORMANCE MANAGEMENT PROMOTE EQUITY?

Tension exists between equity and focus of performance management on efficiency (DeLeon and Denhardt 2000; Radin 2006; Wichowsky and Moynihan 2008). Managerial activities that improve overall organizational performance do not benefit or even worsen traditionally disadvantaged groups such as racial minorities and those with low socioeconomic status (O'Toole and Meier 2004). This tension is seen in several public sectors such as health care, police, social work and is especially pronounced in American public education as performance management is increasingly implemented with the aim of improving students' test scores (Bevan and Hood 2006; Booher-Jennings 2005; Hamilton et al. 2007; Heinrich and Marschke 2010). Studies of test-based accountability policies indicate that although attaching consequences to performance, a part of performance management, can improve measured citizen subgroup performance (Dee and Jacob 2011; Hanushek and Raymond 2005), it can also perpetuate inequity by shifting teachers' efforts from needy citizens not emphasized in

performance assessments to those more likely to improve overall organizational performance outcomes (Booher-Jennings 2005; Hamilton et al. 2007; Neal and Schanzenbach 2010).

Policy scholars have contributed to our collective understanding between test-based accountability, a part of performance management, and citizen subgroup performance (Dee and Jacob 2011; Figlio and Rouse 2006; Hanushek and Raymond 2005). However, their operationalization of performance management has not captured all the tenets of performance management in public management literature. Performance management refers to a system where performance information, created through strategic planning and routine performance measurement, is connected with decision venues and ideally influences them (Moynihan 2008). Performance management in education not only includes attaching consequences to performance, but also goal setting, performance information collection, and managerial empowerment (Moynihan 2008; Snyder, Saultz, and Jacobsen 2017). Managers need not only the information and incentive to focus on results, they also need to be empowered to achieve results (Moynihan 2008; Ouchi and Segal 2003).

Moreover, previous studies rarely take into account both users and providers of public education in their measurement of performance management. Public education is a typical example of technical coproduction where the involvement of service users along with service providers is indispensable for the production of service and outcomes (Osborne, Radnor, and Strokosch 2016). While separate studies have been conducted for performance management for either service users (students) and service providers (teachers and principals) respectively (Fryer 2017), few studies have incorporated both components in modeling. Integrating both components is necessary to fully understand the impact of performance management on educational outcomes

because the outcomes depend on co-production between both service users (students) and service providers (teachers and principals).

This study aims to contribute to the literature by comprehensively operationalizing performance management for both service users (students) and service providers (teachers and principals) and measuring related effects on citizen subgroup performance at U.S public schools. The study directly captures variations in different aspects of performance management ranging from standard-setting, performance monitoring, and attaching consequences for service providers and service users respectively. Since managers needed to be granted autonomy to manage for results, the study also includes the degree of managerial autonomy as part of performance management measurement. This study examines how the variations in the components of school performance management moderate the effect of individual demographic characteristics that are traditionally associated with lower math scores such as racial minority, low socioeconomic status, limited English proficiency, special education status at U.S public schools between 2002 and 2004 (Abedi, Lord and Hofstetter 1998, Barton 2009).

The study found, at most, weak evidence that performance management correlated positively with citizen subgroup performance specified in formal performance evaluation. Although performance management triggered changes in how organizations are run, the results suggest that changes may not be effective in improving measured subgroup performance. The relationships between performance management and unmeasured citizens' performance differ across components. The user-oriented (students) component had positive correlations with underachieving citizens not emphasized in performance assessment while the provider-oriented (teachers and principals) component had negative correlations with the same group of citizens. The findings provide a more nuanced relationship between performance management and

inequity: it can both cause and mitigate the inequity problem of overlooking citizens not emphasized in performance assessment. To address the endogeneity concerns that schools leading the implementation of performance management may be different, I used inverse probability weighting and additional data sets to check the robustness of results. Results from robust checks remained similar.

The rest of the study is organized as follows: the next two sections review theories and evidence on how performance management can promote and impede equity at public organizations in general and at K-12 public schools. Then the data source, measurement, and methodology followed by the presentation and discussion of results. The study concludes with a summary of findings and implications for how performance management can promote equity.

Hypotheses

Although tension exists between performance management and equity, it is also possible to use performance management to promote equity (Kroll 2017). The following section will explain how performance management can advance equity through goal setting, information collection, performance-related rewards/sanctions, and managerial empowerment. Then it proceeds to explain how performance management can cause effort substitution and why effort substitution is problematic and encroaches upon equity in public sector. effort substitution is defined as the practice of reducing effort on non-measured goals and contradicting public values in the process of improving measured performance (Bohte and Meier 2000; Kelman and Friedman 2009; Moynihan 2015).

The Pathway of Performance Management to Promote Equity

Performance management can also be utilized to improve measured performance including the performance of traditionally disadvantaged groups such as racial minorities and citizens from low-income families (Dee and Jacob 2011; Kroll 2017). Performance management

can promote equity in four ways. First, setting standards and goals about specific performance targets can focus attention on the performance and influence decisions on allocating effort and resources in support of achieving the performance targets (Boyne and Chen 2008; Favero, Meier, and O'Toole 2016; Locke and Latham 2002; Poister, Pasha, and Edwards 2013). Hence, goal setting can direct and motivate public service providers and service users to improve the performance of citizen subgroups specified in performance assessment. Similarly, setting standards and goals for student subgroup performance can motivate both service providers and service users to spend more effort on improving student subgroup performance.

Second, collecting performance information and providing comparative feedback on progress in meeting goals relative to historical (how did an organization do compared with itself last year?) and social (how did an organization do compared with its peers?) benchmarks can motivate both service provider and service users to solve problems and maintain strengths in improving student subgroup performance. A variety of studies have suggested that such comparisons are powerful in motivating public employees including teachers, but also the public and politicians (Andersen and Moynihan 2016; Nielsen and Moynihan 2017; Olsen 2017; Taylor and Tyler 2012). In particular, performance feedback draws attention to underperforming units, causing managers to employ strategies to solve problems and enhance workforce productivity, eventually improving organizational performance (Holm 2018; Hong 2019; Rockoff et al. 2012). Frontline workers, such as teachers and service users, can also learn from performance feedback to prioritize their effort and improve performance (Andersen and Nielsen 2019). Related to the pathway, the feedback on teachers' effectiveness, which is based on a combination of experts' subjective evaluations and test scores, can not only inform principals and teachers about the effectiveness of the existing practices but also influence their adjustment of school operation

(Dee and Wyckoff 2017; Rockoff et al. 2012). These changes benefit students in high-poverty schools more than those in low-poverty schools (Dee and Wyckoff 2017). The result implies that students underserved in the previous system are more likely to gain when schools collect and disclose performance information.

Third, attaching rewards and sanctions to performance can incentivize both service users and service providers to improve citizen subgroup performance. Expectancy theory holds that individuals can be motivated toward goals if they see an increase in instrumentality, the belief that accomplishing goals or the failure of doing so will lead to rewards and/or sanctions and valence, the value people place on the rewards of accomplishing goals (Vroom 1964). Attaching consequences to common performance measures for both service users and service providers can enhance both the instrumentality and valence and align their interest in improving the measured performance. Although attaching consequences to performance enhances the importance of goals and induces public managerial' behavior changes in developing countries (Muralidharan 2017), the benefits are largely absent in the United States due to the limited degree of reward and sanctions connected to performance (Bowman 2010; Perry, Engbers, and Jun 2009).

In public education, the effects of attaching consequences to performance on students' subgroup performance have been mixed. While studies examining between-state variation in performance management have found mixed effects of consequential subgroup accountability on racial minorities' test scores (Dee and Jacob 2011; Hanushek and Raymond 2005; Lee and Reeves 2012), studies at the individual level show nonsignificant effects (Springer et al. 2010; 2012). Studies at the school level, the level used in this study, indicate that the findings are contingent on how consequences are attached to performance (Goodman and Turner 2013; Fryer 2017).

Fourth, I added the empowerment of managers as a component of performance management. The tenets of performance management hold that managers needed to be empowered in managing organizational operations to respond in a timely matter to changing circumstances and to achieve results (Osborne and Gaebler 1992; Moynihan 2008). In practice, the empowerment of managers complements high-stakes accountability to realize the potential of improving overall performance (Wang and Yeung 2018; Ouchi and Segal 2003). When empowered managers attain the overall performance goals, they are more likely to have the time and resources to spend time and resources on other goals such as improving citizen subgroup performance (Meier, Favero, and Zhu 2015). In addition, managers need to be granted the autonomy to tailor their responses to meet citizens' heterogeneous needs (Osborne and Gaebler 1992). This is especially true in public education where public schools often consist of students from different racial groups and socioeconomic backgrounds (O'Toole and Meier 2011). Hence, principals at public schools need to be granted autonomy in the management of finances, human resources, and programs to serve different student subgroups' needs (Ouchi and Segal 2003).

Accountability to subgroup performance in NCLB includes students with limited English proficiency, those with special education status, those from low-income families, and racial minorities. These demographic variables have been shown to negatively correlate with student math scores (Abedi, Lord, and Hofstetter 1998; Barton 2009). Therefore, NCLB brings the focus on these traditionally disadvantaged groups to the implementation of performance management.

Hypothesis 1: Performance management will be associated with better performance of citizen subgroups specified in performance evaluation.

The Challenges for Performance Management to Promote Equity

With significant consequences attached to the attainment of flawed goals, performance management can lead to the omission of important values, such as equity. As not every important performance goal can be measured with equal accuracy and certainty, measuring performance and attaching consequences to the performance measures can always prioritize some performance measures at the expense of others (Holmstrom and Milgrom 1991; Holmstrom 2016). The multiple dimensions of organizational performance and accountability to multiple stakeholders in public organizations further aggravate the omission of valuable goals. As public officials are held accountable to these different stakeholders who may champion equity instead of efficiency and effectiveness (Allison 2004; Denhardt and Denhardt 2000), pursuing overall efficiency and effectiveness at the expense of equity is likely to be problematic at public organizations.

Due to the multiple dimensions of performance, the nature of being accountable to multiple stakeholders in the public sector, and the challenge of measuring all the valuable goals equally well, the theoretical model suggests that incentives linked to organizational performance remain weak at public organizations (Dixit 2002). Changing weak incentives to high-stakes performance goals was shown to cause cream skimming in education, health, and job training where citizens who are easy-to-serve or likely to contribute to organizational performance were prioritized while the others were overlooked (Bevan and Hood 2006; Booher-Jennings 2005; Heinrich and Marschke 2010; Tummers et al. 2015).

In public education, a concrete example of prioritizing “high-value” citizens at the expense of others is the practice of favoring students close to performance criteria (Booher-Jennings 2005; Hamilton et al. 2007). A key performance criterion in No Child Left Behind is

whether schools meet state adequate yearly progress (AYP) goals for the percentage of students proficient in reading and math (U.S. Department of Education 2005). The emphasis on the percentage of proficient students incentivized teachers at public schools to spend more time on students close to proficiency level and less time on students far above and below proficiency level (Booher-Jennings 2005; Hamilton et al. 2007; Krieg 2008; Neal and Schanzenbach 2010). As this practice of focusing on students in the middle does not reflect the need of students, especially those low-performing students who need the help most (Jilke and Tummers 2018), this practice detracts from the requirement of equity.

However, some scholars considered the negative effect of performance management on students at the top of score distribution as regression to the mean and showed that the negative effect largely disappeared after controlling for the tendency to return to the mean (Springer 2008). Other studies also found that despite the emphasis on the percentage of students achieving proficiency, students at the bottom of the score distribution and racial minority groups had an increase in their test scores at schools or in states facing the performance pressure (Dee and Jacob 2011; Ladd and Lauen 2010; Reback 2008).

This study investigates the prevalence of focusing on students close to the proficiency level ("bubble kids"; Booher-Jennings, 2005) at the expense of others while controlling for the tendency of regression to the mean. Students in the middle of the score distribution are more likely to be close to the proficiency level than students at the bottom or at the top (Neal and Schanzenbach 2010). In schools effectively implementing accountability for AYP standards, teachers may focus on students whose test scores are close to the proficiency level and, by extension, in the middle of state score distributions because they are more likely to improve school overall performance (Booher-Jennings 2005). Given the limited resources at schools, the

prioritization can be achieved at the expense of other students (Neal and Schanzenbach 2010). The performance management component for service users may further strengthen the signal and discourage struggling citizens from trying to meet performance goals if the chance is slim.

Hypothesis 2: Performance management will benefit citizens who are more likely to contribute to organizational performance more than citizens who are far away from performance criteria and less likely to contribute to organizational performance.

Data

I used the base year (2002) and first follow-up (2004) of the Educational Longitudinal Study of 2002 (ELS:2002) as the study dataset for three reasons. First, ELS:2002 is a nationally representative dataset and the datasets will strengthen the external validity of findings (Ingels et al. 2007). Second, ELS:2002 studied high schools and fill a gap in accountability literature, which usually focuses on primary and middle schools (Dee and Jacob 2011). Third, it is the only dataset to my knowledge that contains archival measures of students' performance and comprehensive measures of performance management that align with performance management tenets (see the following measurement section for details) for both service users (students) and service providers (teachers and principals). High schools were randomly selected first and then students were randomly selected from each participating school. 15360 students out of 750 schools participated in the 2002 survey [1] (response rate was 59.31% at school level and 87.33% at student level). Out of these participants, 11,560 students in 740 schools took the 2004 survey. For every participating student, his or her math and English teacher also filled out surveys. For

[1] To comply with the requirements of using restricted dataset of ELS:2002, all the unweighted sample sizes were rounded to the nearest 10. The presentation of unweighted sample size follows the same rule in the remaining parts of the dissertation.

this study, I limited the sample to the 8,830 students from 570 public schools because all public schools are subject to requirements of school accountability, and the controversy about performance management is heated (Podgursky and Springer 2007).

Compared with schools excluded from the analysis, this sample had fewer Hispanic or African American students, and fewer students participating in special education programs or programs of learning English as a second language (see Appendix A for summary statistics). Respondents had a higher socioeconomic status than nonrespondents. I used sample weights at both school and student level to adjust for the differences between responding and nonresponding schools and students (Ingels et al. 2007).

Measurement

Operationalization of dependent and key independent variables can find support in the literature. I used standardized math scores from ELS:2002 for two reasons. First, compared with the reading scores, the variance of math test scores can be better explained by students' school experiences and school measures (Hanushek and Rivkin 2010). Second, the math questions in ELS:2002 come from three major testing programs: the National Educational Longitudinal Study of 1988 (NELS: 88), National Assessment of Education Progress (NAEP), and Program for International Student Assessment (PISA)(Ingels et al. 2007). As these tests are low-stakes and their scores are less likely to be inflated due to performance pressure (Gigliotti and Sorensen 2020), the scores from these tests are reliable measures of math knowledge and cognitive skills (Hanushek and Woessmann 2010).

The key performance management measures align with tenets in performance management and their applications in public education (Moynihan 2008). In education, performance management involves at least four parts: setting performance standards, a frequent

collection of performance information measuring progress in attaining performance goals, attaching consequences to teachers and students' performance, and frequent disclosure of performance information to ensure public oversight (Snyder, Saultz, and Jacobsen 2017). In addition, public education is a typical example of coproduction. Hence, I measured goal setting, performance monitoring, and attaching consequences to performance for both service users (students) and service providers (teachers and principals) (see Table 1). As managers need to be empowered to manage for results (Moynihan 2008; Ouchi and Segal 2003; Wang and Yeung 2018), a group of variables related to the authority of principals in school management (the empowerment of principals) were also included. The variables within the group above were summed to form six components of performance management (See Table 1). The methodology section explains how I used inverse probability weighting to further alleviate concerns of selection bias of schools to different levels of performance management (Lee and Reeves 2012).

I summed the responses to related questions (see Table 1) to form the components above because the responses included all the measures associated with aspects of performance management. All the variables are positively correlated with the first component and the eigenvalue for the first component is larger than 1. To reduce the dimensions of data, I used principal component analysis on the six components of performance management (Dunteman 1989). I chose the principal component analysis because I did not assume a common latent factor for the performance management measures above and these aspects are often disconnected in practice (Moynihan and Lavertu 2009, Moynihan 2008). Both Kaiser eigenvalue rule and Horn's parallel analysis suggest that two components should be extracted. Before oblique rotation, the eigenvalues for the first and second components are 2.647 and 1.320 respectively. The two

components account for 66.1% of the total variance in the measures of performance management.

Table 1. The Variables Used to Measure Performance Management

Components	Questions
Work planning and standard-setting The response: No (0), Yes (1) The eigenvalue for the first component is 2.062. All four variables have correlations of more than 0.3 with first component. The same number of components as the number of variables were extracted in the analysis. As an underlying factor is not assumed for the variables, Cronbach alpha is not reported.	Whether a school has • Content standards for academic subjects • Content Standard from the State • Content standards linked with performance standards • Competency test tied to content standards
Collection of Students' Performance Data The response: No (0), Yes (1) The eigenvalue for the first component is 2.634. All except two variables had correlations of more than 0.3 with the first component.	Whether math appears on the competency test from grade 7 to 12.
Collection of Teachers' performance Data Response: No (0), Yes(1)	Whether a school has • teachers' evaluation from teachers, • teachers' evaluation from students • teachers' evaluation from principals

The eigenvalue for the first component is 1.304. All but one variables had correlations of more than 0.3 with the first component. For completeness, teachers' evaluation from principals was added in spite of low correlation with the component (0.1).	
Empowerment of Principals Response: No Influence (0), Some Influence (1), Major Influence (2) The eigenvalue for the first component is 3.331. All variables had correlations of more than 0.3 with the first component.	Administrators' perception of • Principal's influence on hiring/firing teachers • Principal's influence on grouping students • Principal's influence on course offerings • Principal's influence on instructional materials • Principal's influence on curricular guidelines • Principal's influence on grading and evaluation • Principal's influence on discipline policies • Principal's influence on school funds
Attaching consequences to teachers' performance Response: No (0) Yes (1). The eigenvalue for the first component is 1.477. All except one variable had a correlation of more than 0.3 with the first component. For compleness of measure, the item of good teachers given priority on requests for materials was included in spite of the relative low correlation (0.22).	Whether administrators hold • Good teachers are given special awards • Good teachers are assigned to better students • Good teachers are given a lighter teaching load • Good teachers are relieved of administrative/disciplinary duties • Good teachers are given priority on requests for materials • Good teachers receive higher pay

Attaching consequences to students' performance Respone: No (0), Available (1), Required (2) The eigenvalue for the first component is 4.883. All variables had correlations of more than 0.3 with the first component.	Whether students need to • Retake competency test if failed • Take a remedial class if fail competency test • Complete competency test preparation class if fail • Take Tutoring/individualized academic program if fail competency test • Take Summer school if fail competency test • Be Referred to alternative/continuing ed school if fail competency test • Pass a test for a high-school diploma

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics.

After the rotation using quartimax criterion to simplify the row of the factor structure matrix (Mulaik 2010), the regression method was used to generate component scores. The first component was more closely related to the performance management component for service users (students); I named it as user-oriented component (UOC) (see Table 2). The second component was more closely related to components for service providers; I named it as provider-oriented component (POC). Both components contain the informational aspect of performance management such as goal setting and collection of information aspect and incentive aspects such as attaching consequences to teachers' and students' performance. The correlation between performance management components for users and providers was 0.3763 suggesting that schools leading the implementation of one component may not necessarily be the leaders in another.

Table 2. The Component Matrix Before and After Oblique Rotation For Study 1

	Before Rotation		After Rotation	
Variable	Comp 1	Comp2	UOC	POC
Teachers' evaluation	0.4707	0.6798	-0.1336	**0.8491**
Setting Standards	0.8634	-0.0623	**0.6876**	0.3847
Consequences to Teachers' performance	0.4072	0.3675	0.0416	**0.5369**
Consequences to Students' performance	0.7927	-0.4536	**0.9137**	-0.0019
Principal's autonomy	0.6547	0.4712	0.1521	**0.7561**
Frequency of Math on annual competency test	0.6759	-0.5398	**0.8881**	-0.1387

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics.

Methodology

To accurately examine the effect of school performance management on individual test performance, three challenges had to be overcome. First, the multilevel nature of the dataset needed to be accounted for, controlling for both individual and school factors that can correlate with test scores. Second, selection biases in implementing performance management had to be addressed. Third, the study must control for the between-state variation in adding the focus on student subgroup performance to performance management. The rest of this section will explain the challenges and attempts to overcome the aforementioned hurdles.

Multilevel modeling is chosen to estimate how performance management in 2002 moderates the association of students' demographic variables with students' test scores in 2004. To get a more accurate estimate, it is necessary to include both individual-level and school-level variables in the model (Raudenbush and Bryk 2002; Snijders and Bosker 2012). Moreover, multilevel modeling enables us to gauge the heterogeneous effects of variables across schools but also more efficiently estimates coefficients of students' demographics (Gelman and Hill 2007;

Raudenbush and Bryk 2002; Snijders and Bosker 2012). Intraclass correlations (ICC) in the full model ranged from 0.24 to 0.31. The ICC result suggests that between-school variance accounts for 24 to 31 percent of the total variance. The large between-school variances (ICC is larger than 0.1) support the usage of multilevel models (Raudenbush and Bryk 2002).

To answer hypothesis one, I created interaction terms between performance management components (TOC and POC) and demographic variables specified in subgroup accountability such as race, limited English proficiency, receiving special education, and low income (socioeconomic status). To answer hypothesis two, Snijder and Bosker (2012) methods of using spline regression to estimate the possible heterogeneous effects of performance management were followed and applied to students above state middle test score in 2002 and those below the average. The model included a state-centered 2002 math score, a quadratic term for math scores from students whose 2002 math scores were below the state middle score. Another quadratic term was created for students whose 2002 math scores were above the state middle scores. Creating the quadratic terms for the two groups respectively can account for the regression to the mean. Interaction terms between performance management component and quadratic terms were created to explore the possibility that performance management may affect students above the criterion differently from those below the criterion. State average was chosen as the primary measure of state middle test scores because proficiency levels on state high-stakes tests, an important performance criterion, can be converted to scores in the same range as the state average math score on the National Assessment of Education Progress (NAEP), the main source of math tests in ELS:2002 (Ingels et al. 2007; McLaughlin et al. 2008; National Center for Educational Statistics 2019). Different operationalizations of the state middle test scores including 40, 50, and 60 percentiles were applied to test the sensitivity of results to different

operationalizations. The spline regression in combination with state-centered math scores in 2002 enabled us to study how students in the middle of score distribution fare compared with those deviating from the state middle. The deviance tests indicated that an independent covariance structure at the school level (level two) and a random slope for the dummy for being in the special education group were statistically significant. Here, the white and Asian students were used as control groups for their relatively high math achievement (Barton 2009).

I controlled for the number of courses students have taken and the teaching experience of the student's math teacher as individual controls because they have been commonly used in studies on students' math test scores (Lee and Bryk 1989). To distinguish between the influence of individual and school-level demographic variables (Raudenbush and Bryk 2002; Snijders and Bosker 2012), the models control for school-level variables including school demographics such as share of male students, shares of African American students, Hispanic students, students receiving special education, and students learning English as a second language and the school average family's socioeconomic scores. Considering the theoretical and statistical significance of these controls (Snijders and Bosker 2012), shares of male students, African American, and Hispanic students, along with school average socioeconomic scores were retained. To ensure the similarity between leaders and laggards in implementing performance management, average school math scores in 2002 were also included (see Table B-1 in Appendix B for the control variables). The following formula describes the final model in the study.

Level One-Students

$$Math2004_{ij} = \beta_{0j} + \beta_{1j} * \left(English\ Second\ Language_{ij}\right) + \beta_{2j} * \left(Special\ Education_{ij}\right) + \beta_{3j} * \left(Black_{ij}\right)$$
$$+\beta_{4j} * (Hispanic_{ij}) + \beta_{5j} * \left(Native_{ij}\right) + +\beta_{6j} * \left(Mixed_{ij}\right) + \beta_{7j} * \left(\ student\ families\ socioeconomic\ score_{ij}\right)$$
$$+\beta_{8j} * (State-centered\_2002\ Math\ score_{ij}) + \beta_{9j} * \left(Centered\_2002\ math\ scores\ below\ \text{state}\ average^2{}_{ij-}\right)$$
$$+\beta_{10j} * (Centered\_2002\ math\ scores\ \ above\ \text{state}\ average^2{}_{ij+}) + \sum_{i=11}^{12} \beta_{ij} * \left(\text{Control}\ Variables_{ij}\right) + r_{ij}$$

Level Two-Schools

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * \left(UOC_j\right) + \gamma_{02} * \left(UOC^2{}_j\right) + \gamma_{03} * \left(POC_j\right) + \gamma_{04} * \left(POC^2{}_j\right) + \sum_{j=5}^{8} \gamma_{0j} * \left(Control\ Variables_j\right) + \mu_{0i}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} * \left(UOC_j\right) + \gamma_{12} * \left(POC_j\right)$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21} * \left(UOC_j\right) + \gamma_{22} * \left(POC_j\right) + \mu_{2i}$$
$$\beta_{3j} = \gamma_{30} + \gamma_{31} * \left(UOC_j\right) + \gamma_{32} * \left(POC_j\right)$$
$$\beta_{4j} = \gamma_{40} + \gamma_{41} * \left(UOC_j\right) + \gamma_{42} * \left(POC_j\right)$$
$$\beta_{5j} = \gamma_{50} + \gamma_{51} * \left(UOC_j\right) + \gamma_{52} * \left(POC_j\right)$$
$$\beta_{6j} = \gamma_{60} + \gamma_{61} * \left(UOC_j\right) + \gamma_{62} * \left(POC_j\right)$$
$$\beta_{7j} = \gamma_{70} + \gamma_{71} * \left(UOC_j\right) + \gamma_{72} * \left(POC_j\right)$$
$$\beta_{8j} = \gamma_{80} + \gamma_{81} * \left(UOC_j\right) + \gamma_{82} * \left(POC_j\right)$$
$$\beta_{9j} = \gamma_{90} + \gamma_{91} * \left(UOC_j\right) + \gamma_{92} * \left(POC_j\right)$$
$$\beta_{10j} = \gamma_{100} + \gamma_{101} * \left(UOC_j\right) + \gamma_{102} * \left(POC_j\right)$$
$$\beta_{ij} = \gamma_{i0}$$

To simplify the presentation of results, I first ran a base model with performance management components, their quadratic terms, and controls to test negative associations of some demographic variables with math scores (Barton 2009). In the second model, the interactions between the performance management components and demographic variables were added to test hypothesis one. In the third model, the interactions between performance management components and state-centered student math score in 2002 were added to the base model to test hypothesis two. The fourth model included all the interaction terms.

To account for possible nonresponse bias, the full multilevel model with survey weights was assessed. To tackle the endogeneity problem, I used the Inverse Probability of Treatment Weighting (IPTW) (Lee and Reeves 2012). Schools with high-performance management scores

can be fundamentally different from schools with low scores. These differences can also account for students' differences in math test scores. For instance, schools in a safe environment and those with more high-achieving students from families with high socioeconomic status may have more stability and flexibility to promote administrative reforms such as performance management (O'Toole and Meier 2011). Schools with more qualified staff or those in states with higher educational expenditures or more experience with implementing performance management in education tend to have the necessary organizational resources to implement reforms including performance management (Bryk et al. 2015). These differences at the school level can also account for the variation in students' math test scores (Krueger 2003; O'Toole and Meier 2011).

To address the selection bias, I used inverse probability weighting to create an apples-to-apples comparison by matching schools leading the implementation of performance management with those lagging behind. Although both matching on and controlling for covariates in regression can tackle selection biases, using a propensity score approach can be advantageous because it explicitly tests whether there is an apple-to-apple comparison by checking whether treatment and control groups share similar values in covariates and propensity scores (Mccaffrey et al. 2013). However, matching can lead to loss of observation while multivariate regression retains all the non-missing observations (Murnane and Willett 2011). For this reason, I tried to combine the strength of both methods and used inverse probability weighting (IPTW).

The TWANG package developed by statisticians at the RAND corporation was used to implement propensity matching. To generate the propensity score, the schools with the bottom one third component score of TOC was coded as zero (the first control group); the schools with middle one third component score were coded as one; the schools with the top one third

component score were coded as two (200, 190 and 190 were assigned to the first, second and third groups respectively). A similar coding scheme was applied for POC (190, 190, and 190 schools were assigned to the first, second, and third group respectively). The inverse of propensity score of a school belonging to its level of performance management was used as the weight in IPTW, so that schools unlikely to be positioned within its performance management level received larger weightings (Ridgeway et al., 2017). Hence, two regressions based on the weights from TOC and POC were run respectively (see Appendix B for details for implementing inverse probability weighting).

ELS: 2002 has shortcomings for answering the question about the effect of performance management on student subgroup performance. As data of performance management in ELS:2002 only covered the starting period of NCLB, it is possible that the dataset may not paint the full picture regarding the implementation of NCLB. NCLB adds a special focus on student subgroups to performance management at public schools. Hence, a lag in implementing the requirements of NCLB can weaken the inference using ELS of 2002 about the effect of performance management on student subgroup performance in two ways. First, states varied in their pace of establishing a disaggregated performance reporting system. The variation means that leaders in school performance management can still overlook student subgroups because their states had not fully implemented the requirement of collecting and disclosing student subgroup performance in the studied period from 2002 to 2004. Second, states may phase in the sanctions for failing to meet the student subgroup goals so that schools have time to adjust to the accountability system. This possibility means that schools with advanced performance management may not experience the incentive to take care of student subgroups during the studied period from 2002 to 2004.

To control for between-state variations in implementing NCLB and connecting performance management with student subgroup performance, four robustness checks were run (see Table B-2 for details of measuring between-state variations). Based on data from *Education Week* (2003, pp. 86–88), I created and controlled for the availability of subgroup performance on school report cards in states as well as the degree to which states apply sanctions in their accountability system to account for the between-state variation in implementing performance management aiming to improve subgroups' performance in the academic year 2002-2003. The second check replaced the sanction index with a measure of both sanctions and rewards. The third test incorporated the length of school accountability until 2002, types of school accountability (none, report-card accountability, and consequential accountability), and state average current educational expenditure per pupil between the academic year 2001-2002 and 2002-2003 years to control for the between-state capacity in implementing accountability to student subgroup performance (Dee and Jacob 2011; Dee, Jacob, and Schwartz 2013; Hanushek and Raymond 2005). The fourth check contained all of the new variables in this paragraph. An additional sensitivity test was also conducted to see whether the result depends on the different operationalizations of state middle score (40, 50, and 60 percentile of students' math scores in 2002 within a state).

Results

The results do not show consistent evidence in support of hypothesis one that performance management will be associated with better performance of citizen subgroups specified in performance evaluation. I first presented the results of multilevel modeling without including any interaction terms to show the association between demographic variables in 2002 and students' math performance in 2004 after accounting for the control variables including school 2002 average math performance. Then, I added the interaction terms between

performance management components and demographic variables in model two and the interaction terms between performance management components and students' relative position in relation to performance standard in model three. Model four included all the interaction terms. The results consistently showed that compared with Whites or Asian students, African American, Hispanic, Native American, mixed-race students, and those learning English as a second language and/or receiving special education had lower math scores in 2004 at five percent significance level (see model 1 in Table3). In the same model, UOC was positively associated with the average of student's math performance. However, neither POC nor UOC consistently moderated the association of demographic variables with student math scores (model 2 in Table 3).

I found mixed evidence in support of hypothesis two that performance management will benefit citizens who are likely to improve organizational performance more than others. As is explained in the hypothesis section, students whose test scores close to performance standard proxied by state average are more likely to contribute to organizational performance and may get higher math performance at the expense of others due to limited resources. The supporting evidence is that the provider-oriented component of performance management in 2002 was positively associated with average student's math performance in 2004, but the quadratic term of the distance between underachieving students' performance score in 2002 and state average score in 2002 negatively moderates the association. For students whose math scores were less than 7 points below the state average in 2002, one standard deviation increase in POC was associated with a significant increase in the students' 2004 math performance. For students whose scores were below the average by more than 17 points, one standard deviation increase in POC was associated with a significant decrease in students' 2004 math performance. By

contrast, the distance variable positively moderated the relationship between the correlation between user-oriented components and math performance. This positive moderating relationship suggests that the further a student is away from the performance standard proxied by the state average, the higher the increase in the student's math score in 2004 is at schools with higher value of user-oriented component of performance management. The findings remained similar in models using either survey weights or inverse propensity scores (see Table 4). The contrasting moderating effect of the distance variable also remained robust in sensitivity tests using 40, 50, 60, and 70 percentile of state scores as the basis for performance standard (results can be obtained upon requests).

The results from robust checks, considering the between-state variation, remained similar as well (see Table C-1 in Appendix C). In addition, the results showed a mixed picture of incentives and information at the state level on improving average students' math performance. Like previous studies (Dee and Jacob 2011; Hanushek and Raymond 2005), the analysis found a consistently positive correlation of consequential accountability before NCLB with student math performance. However, it was also discovered that having in place the sanction measures after NCLB in the 2002-2003 academic year was associated with lower average student math performance, while state implementation of disclosing performance information disaggregated by race, income, English proficiency, and special education status in the 2002-2003 academic year had a significant positive correlation with average student math performance.

Table 3. The Abridged Results of Multilevel Modeling

VARIABLES	1	2	3	4
User-Oriented Component (UOC)	0.393***	0.286*	0.0579	0.0356
	(0.147)	(0.160)	(0.134)	(0.140)
User-Oriented Component^2 (UOC^2)	-0.150	-0.153	-0.110	-0.112
	(0.147)	(0.148)	(0.127)	(0.128)
Provider-Oriented Component (POC)	-0.107	-0.0335	0.308**	0.338**
	(0.137)	(0.160)	(0.127)	(0.137)
Provider-Oriented Component (POC^2)	0.145	0.147	0.129*	0.129*
	(0.0903)	(0.0908)	(0.0773)	(0.0774)
Socioeconomic Score (ses)	3.033***	3.071***	0.571***	0.600***
	(0.157)	(0.162)	(0.0883)	(0.0912)
African American	-5.639***	-5.702***	-0.462**	-0.466**
	(0.369)	(0.386)	(0.207)	(0.218)
Hispanic	-3.498***	-3.620***	-0.0157	-0.0552
	(0.356)	(0.379)	(0.197)	(0.214)
Native American	-3.342***	-3.404***	-0.656	-0.613
	(1.075)	(1.076)	(0.659)	(0.661)
Mixed Race	-1.099**	-1.093**	0.293	0.298
	(0.452)	(0.452)	(0.256)	(0.256)
Learning English as a Second Language	-2.782***	-2.872***	-0.346*	-0.449**
	(0.348)	(0.362)	(0.197)	(0.205)
Receiving Special Education	-6.550***	-6.603***	-1.181***	-1.181***
	(0.381)	(0.389)	(0.237)	(0.243)
UOC* African American		0.199		-0.0512
		(0.343)		(0.213)
POC* African American		-0.0857		0.0503
		(0.324)		(0.203)
UOC* Hispanic		0.247		0.0683
		(0.320)		(0.202)
POC* Hispanic		-0.425		-0.298
		(0.339)		(0.214)
UOC* Native		2.070		-0.596
		(1.407)		(0.823)
POC* Native		-0.157		-0.153

		(1.331)		(0.781)
UOC* Learning English as a Second Language		0.373		0.355
		(0.388)		(0.219)
POC* Learning English as a Second Language		-0.1000		-0.00329
		(0.425)		(0.243)
UOC* Special Education		0.328		0.0259
		(0.424)		(0.265)
POC* Special Education		-0.0443		0.266
		(0.457)		(0.285)
UOC* Socioeconomic Status		-0.180		-0.109
		(0.174)		(0.0978)
POC*Socioeconomic Status		0.243		-0.0448
		(0.184)		(0.104)
State-Centered Math Score in 2002			1.007***	1.007***
			(0.0151)	(0.0151)
The State-Centered Math Score in 2002 Below State Average^2 (Scmathlow)			0.0120***	0.0121***
			(0.000937)	(0.000939)
The State-Centered Math Score in 2002 Above State Average^2 (Scmathhigh)			-0.0102***	-0.0103***
			(0.00108)	(0.00108)
UOC* Scmathlow			0.00198***	0.00176***
			(0.000569)	(0.000607)
POC*Scmathlow			-0.00342***	-0.00352***
			(0.000585)	(0.000624)
UOC* Scmathhigh			0.000475	0.000638
			(0.000641)	(0.000658)
POC*Scmathhigh			-0.000864	-0.000880
			(0.000707)	(0.000730)
Observations	6,950	6,950	6,950	6,950
Number of groups	550	550	550	550
Included controls	Yes	Yes	Yes	Yes
ICC	0.0706	0.0694	0.249	0.252

Note. * 0.1 **0.05 ***0.01 (two-sided tests), and the standard error was clustered at the school level.
Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics. Note. To comply with the requirements of using restricted dataset of ELS:2002, all the unweighted sample sizes were rounded to the nearest 10.

Table 4. The Abridged Results of Weighted Multilevel Modeling

VARIABLES	HLM+Survey Weight	HLM+TOC Weight	HLM+POC Weight
User-Oriented Component (UOC)	-0.386*	0.116	-0.0920
	(0.231)	(0.162)	(0.153)
User-Oriented Component^2 (UOC^2)	0.179	-0.0221	0.00666
	(0.152)	(0.135)	(0.140)
Provider-Oriented Component (POC)	0.385*	0.298*	0.388**
	(0.234)	(0.160)	(0.160)
Provider-Oriented Component (POC^2)	-0.0524	0.0781	0.0687
	(0.129)	(0.0829)	(0.0820)
Socioeconomic Score (ses)	0.562***	0.579***	0.616***
	(0.141)	(0.108)	(0.100)
African American	-0.396*	-0.581***	-0.673***
	(0.239)	(0.219)	(0.223)
Hispanic	0.277	-0.119	-0.274
	(0.270)	(0.247)	(0.225)
Native American	-0.0178	-0.839	-0.719
	(0.746)	(0.811)	(0.883)
Mixed Race	0.00103	0.324	0.362
	(0.345)	(0.319)	(0.385)
Learning English as a Second Language	-0.768**	-0.356	-0.411*
	(0.305)	(0.254)	(0.237)
Receiving Special Education	-1.190***	-1.293***	-1.344***
	(0.360)	(0.242)	(0.239)
UOC* African American	-0.311	-0.270	-0.149
	(0.229)	(0.216)	(0.227)
POC* African American	0.0439	0.256	0.257
	(0.225)	(0.205)	(0.217)
UOC* Hispanic	0.492**	0.0855	-0.0450
	(0.241)	(0.238)	(0.211)
POC* Hispanic	-0.381	-0.262	-0.0290
	(0.272)	(0.305)	(0.255)
UOC* Native	0.674	-0.727	-0.944
	(1.027)	(1.617)	(1.669)

POC* Native	-1.674**	0.0222	0.486
	(0.732)	(1.052)	(1.107)
UOC* Learning English as a Second Language	0.435	0.209	0.378
	(0.275)	(0.266)	(0.296)
POC* Learning English as a Second Language	-0.0394	0.290	0.243
	(0.385)	(0.280)	(0.284)
UOC* Special Education	-0.369	0.0861	-0.105
	(0.432)	(0.327)	(0.312)
POC* Special Education	0.116	0.136	0.355
	(0.460)	(0.289)	(0.279)
UOC* Socioeconomic Status	-0.0969	-0.0696	-0.185
	(0.152)	(0.117)	(0.114)
POC*Socioeconomic Status	0.0617	-0.0908	-0.0897
	(0.146)	(0.120)	(0.112)
State-Centered Math Score in 2002	0.998***	1.002***	0.994***
	(0.0244)	(0.0194)	(0.0199)
The State-Centered Math Score in 2002 Below State Average^2 (Scmathlow)	0.0113***	0.0112***	0.0105***
	(0.00141)	(0.00119)	(0.00111)
The State-Centered Math Score in 2002 Above State Average^2 (Scmathhigh)	-0.0102***	-0.00960***	-0.00864***
	(0.00183)	(0.00129)	(0.00141)
UOC* Scmathlow	0.00220*	0.00202**	0.00265***
	(0.00114)	(0.000876)	(0.000759)
POC*Scmathlow	-0.00243*	-0.00377***	-0.00450***
	(0.00129)	(0.000921)	(0.000853)
UOC* Scmathhigh	0.00120	0.000540	0.000832
	(0.000926)	(0.000776)	(0.000824)
POC*Scmathhigh	-0.00189	-0.000640	-0.000930
	(0.00115)	(0.000906)	(0.000996)
Observations	6,946	6,946	6,946
Number of groups	548	548	548
N	6946	6946	6946
ICC	0.274	0.309	0.270

Data Source: Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics.
Note. * 0.1 **0.05 ***0.01 (two-sided tests), and the standard error was clustered at school level. To comply with the requirements of using the restricted

dataset of ELS:2002, all the unweighted sample sizes were rounded to the nearest 10.

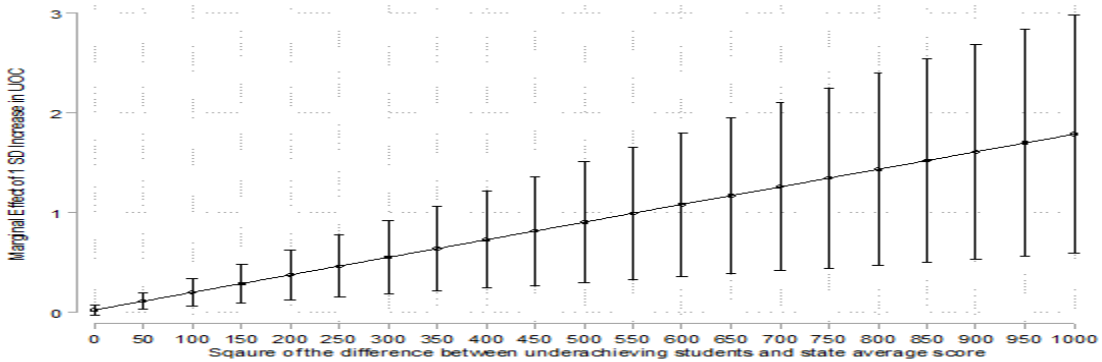

Figure 1.Marginal Effect of 1 SD increase in POC from the Mean for Underachieving Students

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics.

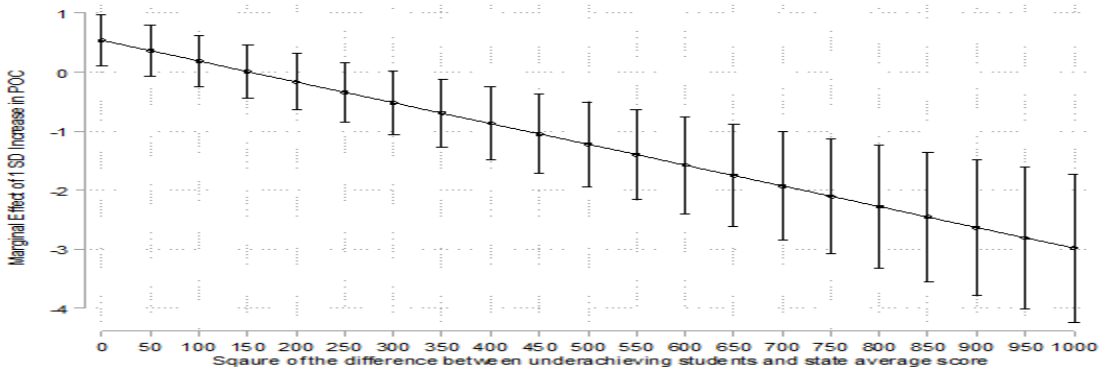

Figure 2. Marginal Effect of 1 SD increase in UOC from the Mean for Underachieving Students

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics.

Discussion

To recap, I found an unfulfilled promise of performance management for promoting equity. NCLB had added a focus on student subgroups' performance to school accountability. Despite evidence of positive association of UOC and POC with average student academic performance, neither component significantly reduced the academic achievement gap between advantaged groups (White and Asian) and disadvantaged groups specified in NCLB. In addition, I found evidence in support of previous literature that performance management can cause inequity in public education (Booher-Jennings 2005). At schools leading the implementation of POC in 2002, students close to the performance standard had higher math score in 2004 while students whose math performance in 2002 was far below the standard had lower math score in 2004. Still, some performance management components can mitigate the tension between efficiency and equity. At schools leading the implementation of UOC in 2002, the same group of underachieving students got higher math scores in 2004.

One way to understand the unfulfilled promise of performance management for equity is to revisit the underlying theories of changes for performance management. Through goal setting, performance tracking, attaching consequences to performance, and granting managerial autonomy, performance management aims to identify problems and spur changes that eventually solve the problems and improve performance. Like previous studies (Massell et al. 2005; Hamilton et al. 2007), I found that performance management was associated with organizational changes to overcome the hurdles for improving the academic performance of students, especially those from traditionally disadvantaged subgroups. The organizational changes range from the implementation of block schedules for academic subjects to increasing the percentage of students receiving remedial math courses and percentages of

teachers receiving more training on teaching to students with limited English proficiency (see Table C-1 in Appendix C). A systemic review of the literature indicates that increasing the amount of time spent on instruction increases average academic performance; although teachers' professional development increases their knowledge of teaching practices, it does not significantly improve student academic performance (Fryer 2017). Hence, it is possible that although implementing performance management generates changes to improve subgroup academic performance specified in the accountability regime of NCLB, the changes such as professional development on teaching to disadvantaged groups were ineffective in improving measured student subgroup performance.

Attributing the null effect to the ineffectiveness of training is consistent with the research on public education and training on performance management in government. The lack of know-how to implement effective interventions in various settings, including conducting training for teaching to diverse student subgroups, is a common challenge in public education (Bryk et al. 2015). Given students' heterogeneous needs and limited resources and time for teachers and principals in public schools (Ouchi and Segal 2003), the process of translating input (trainings) to desired outcomes (effective teaching that improves student subgroups' test scores) is still largely a black box. In addition, training offered to employees at both federal and state governments positively correlated with implementation and convergence of performance management measures (Kroll and Moynihan 2015; Gerrish and Spreen 2017), but it was not associated with better average performance (Spreen, Afonso, and Gerrish 2020). One possible reason is that training has not improved the public employees' capacity to overcome the hurdles in performance improvement (Kroll and Moynihan 2015).

Revisiting the pathways from performance management to equity may also illuminate why UOC and POC correlated differently with underachieving student academic performance. To better serve struggling citizens, performance management must

have accurate performance information to identify areas of improvement. UOC can be better than POC in providing accurate performance feedback. UOC relies on competency tests and POC on teacher evaluations to track performance information. While competency tests contain valid information of student and teachers' performance teachers and principals can learn from to improve performance (Andersen and Nielsen 2019), teacher evaluation often fail to distinguish between outstanding and mediocre teachers (Weisberg et al. 2009). The difference means that UOC had more accurate information to identify who need additional assistance and inform the decision process to improve performance (Rockoff et al. 2012). Hence, UOC may enable schools to identify and assist struggling students and counter the prioritization of students close to performance standards while POC may reinforce the prioritization due to lack of alternative and accurate prioritization criteria.

Moreover, the connection between performance and consequences in UOC may be better implemented to encourage public employees to take care of struggling citizens in need of help. Although implementing the measures of attaching consequences to performance for both users and providers in Table 1 can generate a supportive social environment within public organizations, Table A-1 indicates the measures were better implemented for users than for service providers. On average, schools have implemented several measures of linking consequences to students' performance while having adopted less than one measure of connecting teacher performance with consequences. This means that schools with a high POC score still have not implemented many measures of recognizing and supporting outstanding teachers to keep up with the good work while schools with high UOC scores have offered various assistance to students who fail competency exams, thereby indirectly assisting in students' teachers' work of improving student outcomes. Offering organizational support to public employees can foster a sense of supportive

environment within public organizations (Corduneanu, Dudau, and Kominis 2020). A supportive social context can prevent the encroachment of performance management upon public employees' public service motivation. With public service motivation intact, public employees may follow their internal call of taking care of needy citizens such as struggling students even though doing so may not maximize organizational performance in performance assessment (Jilke and Tummers 2018).

The findings hold several implications for studies on performance management. First, it shows a nuanced picture between performance management and inequity and the value of incorporating performance management components for service users and service providers. Performance management can both cause and mitigate inequity. Consistent with previous literature (Neal and Schanzenbach 2010; Radin 2006), the study also found that the provider-oriented component of performance management advanced efficiency at the expense of equity. However, it also found that the user-oriented component of performance management mitigated the trade-off by improving struggling citizens' performance.

Second, the nuanced relationship between performance management and inequity can illuminate how performance management can mitigate inequity in the future. Performance management can mitigate inequity if it (1) provides accurate performance feedback to identify areas of improvement and citizens needing help; and (2) creates a more supportive environment when applying consequences to performance. The recommendation is also in line with the recent discussion on how satisfying basic needs and addressing social contexts may mediate the relationship between performance management and public service motivation (Corduneanu, Dudau, and Kominis 2020). It has often been claimed that performance management crowds out public service

motivation because the connection between consequences and performance may limit individual autonomy. However, performance management can also satisfy the basic needs for competency and autonomy. For instance, performance management can generate accurate and timely performance information. Timely feedback can contribute to the learning and improvement process, thus leading to a feeling of competency. To reduce the potentially detrimental effects of attaching consequences to performance onto perceived autonomy, performance management can utilize non-financial incentives such as training, fewer administrative burdens, and more flexible schedules to foster a perception of a supportive environment at public organizations. Satisfying these basic needs may, in turn, maintain strong public service motivation to serve all citizens regardless of formal incentives.

A limitation of this study is the lack of exogenous variation in performance management. The choice of research design reflects an inherent trade-off between internal and external validity. Given the large variation in local implementation of performance management and the lack of studies to capture all the variations (Dee and Jacob 2011; Massell et al. 2005), I chose the study design to capture the wide variation in components of performance management across different settings at the expense of exogenous variation in performance management. Capturing this variation enables us to discover different associations of performance management components with citizens unlikely to contribute to organizational performance. A large-scale exogenous change for future studies is the enhancement of teacher evaluation and linking evaluation results to decision-making after the Obama administration incentivized states to adopt these measures to increase their chances to win a Race to the Top grant.

Conclusion

Performance management, in theory, can improve citizen subgroups' performance by providing information and incentives on improvement in disaggregated performance indicators. This study of a sample of nationally representative high schools suggests that the potential of performance management for promoting equity is largely elusive in public education. Although performance management indeed triggers changes aiming to improve both average and citizen subgroups' performance, the results suggest that process changes were not effective in attaining the goal of improving test scores of traditionally disadvantaged groups. The findings confirmed that performance management can change the operation of public organizations (Hamilton et al. 2007; Moynihan 2008), but the changes can be of limited value to achieve the desired organizational outcomes when the know-how from process changes to the desired outcomes is lacking (Bryk et al. 2015).

The study also shows a nuanced relationship between performance management and inequity. Like previous studies (Booher-Jennings 2005; Heinrich and Marschke 2010), the study found that performance management can impede equity. The study shows a negative correlation of the provider-oriented component (POC) of performance management with citizen subgroups with the limited potential of improving organizational performance. However, a comprehensive measurement of performance management enables us to find a positive correlation of user-oriented components of (UOC) performance management with the same citizen subgroups. The contrast suggests that performance management can also mitigate inequity created in the process of pursuing efficiency.

STUDY 2: DOES PERFORMANCE MANAGEMENT DEVELOP A SKILLED WORKFORCE AT THE EXPENSE OF DEMOCRATIC CITIZENSHIP

Performance management has been widely applied in the public sector for its promise to deliver improvement in key performance metrics (Moynihan and Beazley 2016). Its widespread application also raises the concern of effort substitution where laser-focused efforts to attain measured and high-stakes performance targets marginalize important but low-stakes or unmeasured performance outcomes (Kelman and Friedman 2009). A concrete example of the effort substitution concern is the claim that performance management in public education directs attention toward workforce preparation at the expense of democratic citizenship. The alleged detrimental effect of performance management on citizenship warrants attention because it conflicts with the government's responsibility to develop active citizens, a critical element of effective governance in diverse areas (Bryson, Crosby, and Bloomberg 2014; Musso, Young, and Thom 2019). Despite the importance of this topic, few empirical studies exist. This study fills this gap and examines how performance management affects civic participation through its effects on students' capacity for and attitudes towards civic participation at American public high schools. Civic participation is defined as "working to make a difference in the civic life of our communities... It means promoting the quality of life in a community, through both political and non-political processes" (Ehrlich 2000, vi).

Since the 1983 publication of *A Nation At Risk,* depicting failing public schools, the U.S. has witnessed a gradual shift to a new paradigm emphasizing the economic significance of education (Mehta 2013). Comparative performance metrics showing those American students lagging behind peers were given additional weight because they were assumed to reflect slipping economic

competitiveness. Test scores became the de facto measure of progress on this dimension. This new paradigm culminated in the enactment of the 2001 No Child Left Behind Act (NCLB), a nationwide test-based accountability system. Implementing test-based accountability put performance management in the center of school management, representing an example of a policy change that altered administrative practices (Moynihan and Soss 2014).

A classic problem of performance management systems is effort substitution, with unmeasured goals being overlooked relative to those captured in formal indicators (Kelman and Friedman 2009). As Chester Finn and Diane Ravitch (2007), who had been early advocates of NCLB and Assistant Secretaries of Education, observed, "NCLB puts a premium on reading and math skills…Worthy though these skills are, they ignore at least half of what has long been regarded as a 'well rounded' education…" Hence, implementing a performance management revolving around test scores can crowd out activities that foster civic skills and civic norms, thereby enhancing civic participation. Civic skills refer to "the communications and organizational abilities that allow citizens to use time and money effectively in political life" (Verba, Schlozman, and Brady 1995, 304). Civic norms refer to "a shared set of expectations about the citizen's role in politics" (Dalton 2008, 78).

The purported negative effects of performance management on civic participation may have severe long-term consequences for governance. Civic participation is a habit formed early in life (McFarland and Thomas 2006) and can drive engagement with the government and improve government performance in the future (de Geus et al. 2020). Hence, the negative effect of performance

management on adolescents' civic participation may not only impede government performance improvement but also defeat the purpose of developing trust in government through performance management (Wichowsky and Moynihan 2008).

However, the alleged negative effects of performance management on civic participation may not tell the whole story. The focus of performance management on test scores and economic skills may have some positive effects on people's ability to participate in public affairs and serve the community. Resource models for civic participation predict that people with strong civic skills and high cognitive abilities will find participation easier and will tend to participate at higher levels (Musick and Wilson 2008; Verba, Schlozman, and Brady 1995). By improving students' reading and math scores (Dee and Jacob 2011; Favero, Meier, and O'Toole 2016), performance management can improve civic participation through its positive effects on cognitive abilities. For this reason, an assessment focusing only on the potentially negative impact of performance management on civic participation risks being incomplete.

Previous research shows that despite the concern that performance management undermines civic participation (R. Rothstein, Jacobsen, and Wilder 2008), performance management is found to promote civic participation by making public officials value more citizens' inputs (Kroll, Neshkova, and Pandey 2019). However, few have empirically examined how performance management shapes citizens' ability and attitudes for participation. In education, previous studies have mostly examined unintended consequences of performance management on academic instructions to improve students' economic skills (Booher-Jennings 2005; Figlio and Rouse 2006; Jacob 2005; Neal and Schanzenbach 2010). Few studies have systemically examined the consequences of performance

management on the other educational goals such as fostering citizenship, even though attaining these goals is supposed to be a defining characteristic of public organizations.

This study aims to fill the gap and reconcile the mixed effects of performance management on civic participation. Using path analysis on a nationally representative sample of American public high school students from 2002 to 2006, I simultaneously examined the indirect effects of performance management on civic participation through its effects on cognitive abilities, participation in activities developing students' civic skills, and students' commitment to civic norms. As public education epitomizes coproduction (S. P. Osborne, Radnor, and Strokosch 2016), I organized performance management components by the division of users (students) and providers (teachers and principals). The findings showed that the user-oriented component of performance management had a consistently positive but small indirect effect on civic participation by improving students' cognitive abilities while the provider-oriented component had a consistently negative but small indirect effect by reducing students' participation in extracurricular activities where they developed civic skills. Performance management did not shape students' commitment to civic norms.

The remainder of this study is organized as follows. The next section sets up the framework for selecting variables influencing civic participation. It then draws on policy feedback theory to build the study hypotheses. The third section describes data source. The fourth section introduces measures of key variables. The fifth and sixth sections present methods and findings. The seventh section discusses the implications of findings. The conclusion closes by summarizing findings and limitations and suggesting directions for future research.

Theories of Civic Participation

I drew on Verba, Schlozman, and Brady's (1995) and Wilson's (2000) frameworks in organizing the theories accounting for the individual level of civic participation. The theories can be classified into three types. The first type holds that people are more likely to engage in civic participation if the benefits of civic participation exceed its cost (Downs 1957; Klandermans 1984; Olson 1971; Verba, Schlozman, and Brady 1995). The second type argues that people tend to participate if they are convinced of the importance of civic participation (Dalton 2008). The first two schools of thought assume that people make the decision on civic participation in isolation while the third one assumes that the environment, especially social network, plays an important role in shaping individual behaviors of civic participation (Wilson 2000). The third type of theories explores how the depth, width, and composition of social network affect individual willingness and opportunities for participation and eventually their likelihood of civic participation (Bond et al. 2012; Mutz 2002; Rosenstone and Hansen 1993).

Civic Participation as a Rational Choice

The key idea of rational choice theory for civic participation is that people weigh the costs and benefits of civic participation and will only participate if the benefits exceed the costs (Verba, Schlozman, and Brady 1995). Expectancy theory holds that the benefits of civic participation depend on how much an individual values the rewards from a specific outcome of participation and how likely his or her participation contributes to the achievement of the outcome (Klandermans 1984). If the rewards from the outcome are limited, the increase in the group size will decrease the value an individual assigns to an outcome and the possibility that his participation increases the likelihood of goal attainment (Olson 1971). The early literature focuses on the rewards of participation such

as selective material benefits and changes in collective outcomes (Olson 1971). It was criticized for having a low predictive power regarding actual participation (Verba, Schlozman, and Brady 1995).

To compensate for the low predictive power of focusing on the individual benefits, scholars emphasized individual abilities to meet the costs of participation (Verba, Schlozman, and Brady 1995). The civic voluntarism model holds that individual differences in time, money, knowledge, and civic skills will limit their ability to meet the costs of participation and shape both the level and mode of participation (Verba, Schlozman, and Brady 1995). Possessing in-depth knowledge about civic participation and related skills are the key to ensuring a high level of civic participation.

Along with human capital theory, the civic voluntarism model holds that education boosts civic participation in three ways (Musick 2008; Verba, Schlozman, and Brady 1995). First, education provides people with the knowledge necessary for civic participation. Civic participation often requires a mixture of knowledge including economics, government, laws, history, and even science. The longer students stay in school, the more related knowledge they will learn (Rosenstone and Hansen 1993). Second, formal education also helps students develop skills they can apply in civic participation. Formal education trains students to handle the common tasks in civic participation such as the ability to give speeches, read, and write letters (Verba, Schlozman and Brady 1995). They have developed skills to learn what they need to know in civic participation such as the ability to unpack abstract and complex ideas and search and evaluate information (Campbell 2006; Rosenstone and Hansen 1993). The improvements in both knowledge and skills will lower the cost of civic participation and make people more comfortable with and confident in participation (Condon 2015).

Third, the civic voluntarism model holds that education attainment will affect the type of jobs after education and opportunities for exercising civic skills at work (See Verba, Scholzman, and Brady [1995] for a detailed discussion).

The civic voluntarism model cannot fully explain the puzzle of why voter turnout (excluding the anomaly of the 2020 US Presidential Election) has failed to increase with the rising level of absolute educational attainment. Some studies argue that educational attainment was just a proxy of variables such as family socioeconomic status that affects both participation and educational attainment (Berinsky and Lenz 2011; Kam and Palmer 2008; Tenn 2007). However, other studies also reaffirmed the positive effect of individual educational attainment on participation (Dee 2004; Milligan, Moretti, and Oreopoulos 2004; Sondheimer and Green 2010). Some scholars hold that relative educational attainment, which signals relative social status, rather than the absolute level of educational attainment significantly affected civic participation (Nie, Junn, and Stehlik-Barry 1996), but evidence for this argument is mixed (Campbell 2009). Another explanation is that educational quality, which is measured by test scores, matters more than educational attainment for participation (Condon 2015).

Civic Participation as a Duty

Rational choice theory accounts for both the benefits and costs of civic participation. Contrasted with other theories, it offers a better explanation regarding who will participate, but it is not sufficient to account for the variation in civic participation for two reasons. First, it has not offered a satisfactory explanation for types of civic participation that require relatively little time and effort, such as voting. The number of opportunities for exercising civic skills was found to not correlate with the possibility of voting (Verba,

Schlozman, and Brady 1995). Second, it does not reveal a full picture of why people are motivated to participate in activities that require more investment of time and effort such as volunteering.

The values and norms in favor of helping others in need, community, and society at large can further supplement the explanations for the level of civic participation. To hold together a diverse society where reason and common bonds are fragile, John Dewey (1916) held that schools should instill a set of democratic values in students. To instill democratic values, education should arouse interest in social relationships and cultivate a habit of mind that makes changes without causing social disorder (Dewey 1916, 115). From this vantage point, liberal democracy is primarily a way of associated living that will make society more participatory and equal. Hence, forging and promoting the value of liberal democracy can boost civic participation (Rockefeller 1993).

That said, values can be vague and used to justify different forms of behaviors. Hence, it is necessary to identify the values and their connection with behaviors to make the case that values induce civic participation (Musick and Wilson 2008). The values include altruism and civic duty and are rooted in humanitarian and religious values. These values suggest that it is one's obligation to help those less fortunate and contribute to bettering the community and society. People endorsing altruism as a value are more likely to volunteer in the next period (Musick and Wilson 2008). Volunteering refers to activities where labor is provided for free to help a person, a group or a cause (Wilson 2000). These values are reflected in the norm of generalized reciprocity and the norm of social responsibility. Norms are "rules of conduct" identifying a course of action in specific situations (Musick and Wilson 2008, 97). Parents can facilitate children's endorsement of social responsibilities and generalized reciprocity by not only preaching but practicing

these values (Wilson 2000; Musick and Wilson 2008). In addition, children are more likely to engage in volunteering behaviors if their parents provide a supportive and cohesive environment (Eisenberg et al. 1992; Lawford et al. 2005; Rosenthal, Feiring, and Lewis 1998).

Civic Participation as an Interaction

The theories in the previous two sections focus on the variation in the individual willingness and ability to participate. The theories in this section account for the differences in people's opportunities for participation and their likelihood to take the opportunities. In other words, the previous two sections focus on the supply of civic participation effort and assume that people make participation decisions on their own. However, organizations are not passive recipients of individual participation effort (Verba, Schlozman, and Brady 1995; Wilson 2000; Musick and Wilson 2008). They reach out to individuals with particular demographic characteristics and social capital. The variables related to social network were used as control variables. Hence, I did not review that literature here. Interested readers can read Musick and Wilson's (2008) and Verba, Schlozman, and Brady's (1995) books for detailed discussions on the topic.

The Indirect Effects of Performance Management on Civic Participation

Performance management can have indirect effects on civic participation through its effects on students' resources related to civic participation such as (1) cognitive abilities, (2) civic skills, and (3) attitudes toward civic participation. First, performance management is anticipated to have a positive indirect effect on civic participation through its positive impact on cognitive abilities. Standardized test scores have been established to be an appropriate measure of students' cognitive abilities (Hanushek and

Woessmann 2008). Performance management can improve average reading and math scores by setting clear and challenging goals (Boyne and Chen 2008; Favero, Meier, and O'Toole 2016), prompting managers to adjust strategies according to feedback (Holm 2018), attaching consequences to performance (Dee and Wyckoff 2015; Taylor and Tyler 2012), and empowering managers to manage for results (Wang and Yeung 2018). The improvement in cognitive abilities can reduce the costs of participation and increase students' civic participation (Verba, Schlozman, and Brady 1995).

Hypothesis 1 (H1): Performance management will increase civic participation by improving students reading and math abilities.

Effort substitution is the logic underlying the following two hypotheses. As multiple measured tasks performed by an agent cannot align perfectly with what principals care about, a mismatch exists between what contributes to organizational mission and activities organizations reward. For this reason, paying for performance targets can lead to the omission, duplication, or null effect of agents' efforts for important goals (Holmstrom and Milgrom 1991). Previous studies have shown at least three aspects of effort substitution in academic instruction at public schools. First, the reliance on high-stakes tests makes teachers teach to the tests. As a result, performance improvement on low-stake questions/tests is less significant than on high-stakes questions/tests (Figlio and Getzler 2006; Jacob 2005). Second, the utilization of cut-off points as the major accountability criterion leads to the prioritization of students close to the cut-off points, sometimes at the expense of others (Booher-Jennings 2005; Neal and Schanzenbach 2010). Third, schools allocate resources to high-stakes subjects away from low-stake subjects such as science and social studies (Dee, Jacob, and Schwartz 2013). All the examples examine the effort substitutions in pursuing the goal of improving students' economic skills. None of the

studies, to our knowledge, have examined the effect of performance management on effort substitution among different educational goals.

Developing a citizenship norm and civic skills are valued educational goals (R. Rothstein, Jacobsen, and Wilder 2008). The high stakes attached to achievement in reading and math can cause the goal of enhancing civic participation to be overlooked because high-performance pressure narrows teachers' and students' attention to high-stakes performance goals. Participating in extracurricular activities has been consistently found to improve civic participation. Longitudinal studies showed that adolescent participation in extracurricular activities correlated positively with adult civic participation (McFarland and Thomas 2006). Participating in extracurricular activities can enhance civic participation in two ways. First, participating in extracurricular activities provides opportunities for students to practice civic skills. Second, extracurricular activities can improve interactions among students and raise students' sense of belonging (Feldman and Matjasko 2005). Such involvement can be conducive to the development of a sense of the common good and of moral and ethical responsibilities to others, which, in turn, predisposes students favorably toward civic participation (Musick and Wilson 2008). The emphasis on math and reading test scores lead to teachers and students spending more time on test preparation and less time on activities not directly related to test taking, such as extracurricular activities.

Hypothesis 2 (H2): Performance management will reduce civic participation by reducing students' participation in extracurricular activities.

Performance management can change students' attitudes toward civic participation by shaping their day-to-day experiences. Students' experiences with school institutions and teachers can be as potent as, if not more than, traditional education based on the transmission of curriculum knowledge in shaping students' attitudes towards civic participation (Dewey 1937). In interactions with teachers and staff, the street-level bureaucrats at public schools, students (citizens) have first-hand experiences of the expectations of state (Lipsky 1980; Soss 1999). The experiences can shape students' attitudes toward future civic participation. If performance management "teaches" students lessons related to test scores and civic participation, students may conclude that test scores are of primary importance and that civic norms are of secondary value. The slipping commitment to civic norms will lead to lower civic participation.

Hypothesis 3 (H3): Performance management will reduce civic participation by diminishing students' commitment to civic norms.

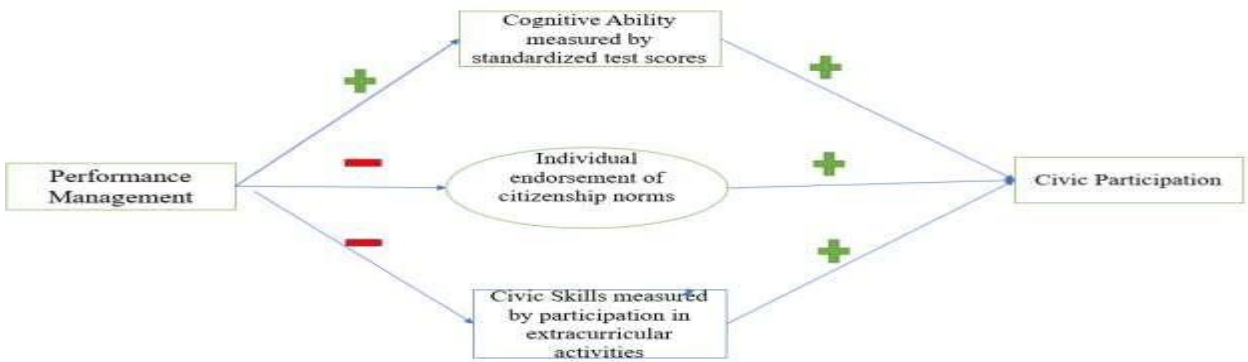

Figure 3. Hypothesized Relationships Between Performance Management and Civic Participation

Source: Original Figure Created for the Dissertation.

Data

Most of the data are from the base year (2002) and first follow-up (2004) of the Educational Longitudinal Study of 2002 (ELS:2002). To increase internal validity and reduce bias, robust checks also used student volunteering behaviors from the second follow-up (2006) of ELS: 2002 while controlling for the volunteering behaviors in the base period (2002). Common Core of Data (CCD) from the academic year 2001-2002 is the source for school demographic information. CCD is based on the administrative data state educational agencies provide to the Department of Education and gives descriptive statistics of school basic characteristics. Although no every school in ELS: 2002 could be matched with the record from CCD, only a small proportion of students were at public schools without CCD data (500 out of 8,830.).

For ELS: 2002, high schools across the nation were randomly selected. Then, students in tenth grade in 2002 were randomly selected from each participating school. The response rate was 59.31% at the school administrator level and 87.33% at the student level (Ingels et al. 2007). I used students' responses to construct student level measures and administrators' answers to create school level measures. I limited the sample to public schools (8,830 students from 570 schools) because all public schools are subject to requirements of school accountability and the controversy about performance management is heated (Podgursky and Springer 2007). Compared with observations in ELS:2002 but excluded from the analysis, this sample had fewer Hispanic or African American students. Respondents had a higher socioeconomic status than non-respondents. Since the shares of minority students have been used as measures of task difficulty (O'Toole and Meier 2011), the analysis based on the sample may overestimate the effect of performance management on students' civic participation.

Measurement

This section started with the description of how performance management, the key independent variable, and civic participation, the key dependent variables, were operationalized. Key intermediary variables related to civic participation were introduced in the order of their relations with cognitive abilities, civic skills, and citizenship norms (details of these variables can be found in Appendix D). The methodology section will briefly explain rationales for including control variables (details of control variables can be founded in Appendix E).

Measuring Performance Management

In education, performance management involves at least four parts: setting performance standards, frequent collection of performance information measuring progress toward attaining performance goals, attaching consequences to teachers' and students' performance, and frequent disclosure of performance information to ensure public oversight (Snyder, Saultz, and Jacobsen 2017). I utilized the same measure of performance management as in study one (above) except that the frequency measures of English testing were also added.

The summary statistics of the six components (see Appendix F) indicated uneven progress in implementing different components of performance management in 2002. Most schools set goals and measured progress towards their goals. More than half of schools set standards for academic subjects, linked them with tests and performance standards, used at least one form of teacher evaluation, and tested math at least once in the annual competency test. In contrast, progress linking consequences to performance is

limited. Although at least half of schools had attached consequences to students' performance to a medium degree, an equally large number of schools did not adopt a single survey measure for rewarding outstanding teachers. Most principals enjoyed sizeable discretionary power in the management of their schools. The description of performance management implementation is consistent with the previous studies (O'Toole and Meier 2011; Moynihan 2008; Weisberg et al. 2009).

To reduce the dimensions of data, principal component analysis (PCA) was conducted. To avoid skewing the results, all the variables used in the PCA were first standardized. Then PCA was run on their correlation matrix. PCA was chosen because a common latent factor for the performance management measures was not assumed due to the disconnections between these components in reality (Moynihan 2008; Moynihan and Lavertu 2012). Both Kaiser eigenvalue rule and Horn's parallel analysis suggested that two components should be extracted. The two components accounted for 68 percent of the variation. After the oblique rotation which allows for correlation between components, the first component was more closely related to subcomponents for students like the frequency of tests and attaching consequences to student performance. Hence, the component was named as a user-oriented component (UOC) of performance management. The second component was more closely related to subcomponents for public employees at schools, such as—teacher evaluation, principal autonomy, and attaching consequences to teacher performance. For this reason, the component was named as a provider-oriented component (POC) of performance management. Setting standards correlated with both components to a medium degree since standards will affect both service users and providers. The two components only weakly correlated with each other (Pearson correlation being 0.219), therefore, each contains performance management information unique to users and providers respectively.

Table 5. The Component Matrix Before and After Oblique Rotation for Study 2

	Before Rotation		**After Rotation**	
Variables	Comp1	Comp2	UOC	POC
Teachers' evaluation	0.3245	0.7355	-0.159	**0.8236**
Setting standards	0.8023	0.2124	**0.5494**	**0.5134**
Consequences to teachers' performance	0.3333	0.4102	0.0393	**0.5186**
Consequences to students' performance	0.8362	-0.2004	**0.8201**	0.1352
Principal's autonomy	0.5191	0.6254	0.0691	**0.7948**
Frequency of Math on annual competency test	0.8327	-0.425	**0.949**	-0.0792
Frequency of English on annual competency test	0.8327	-0.4191	**0.9455**	-0.0735

Data Source: Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics.

Measuring Civic Participation

Student participation in volunteering activities in 2004, the number of volunteering categories in 2004, and the frequency of community service in 2004 were used as measures of civic participation. Individual adolescent volunteering is important for at least two reasons. First, volunteering can encourage social conformity and reduce teenage delinquency (Allen and Philliber 2001). Second, adolescent volunteering has a persistent positive impact on adult civic values and behaviors (Astin, Sax, and Avalos 1999; McFarland and Thomas 2006). The same volunteering measures in 2006 were also used in the robust checks to mitigate the concern of reverse causality. A meaningful variation in volunteering behaviors exists in the studied period because only the District of Columbia implemented a community service requirement for graduation at the State level (National Center for Educational

Statistics 2006).

The summary statistics (see Appendix F) showed that the participation rate in volunteering at sampled public high schools increased from 31 percent in 2002 to 60 percent in 2004 before declining to 45 percent in 2006. The frequency and the type of volunteering followed a similar pattern, increasing from 2002 to 2004 when students graduated from high school while declining from 2004 to 2006. These patterns confirm the life cycle of volunteering: volunteering activities for young people tend to peak during high school years (Musick and Wilson 2008). Compared with the narrow categories of volunteering, students' participation in extracurricular activities was broader. At least 50 percent of students participated in two or more extracurricular activities in both 2002 and 2004. In addition, more than 50 percent of students spent at least one to four hours on extracurricular activities every week in 2004.

Intermediary Variables Related to Civic Participation

To measure students' cognitive abilities to process information and make decisions related to civic participation, I utilized standardized math and reading scores in ELS:2002 (a mean of 50 and a standard deviation of 10). Using the participation in the activities offering opportunities to practice civic skills as measures of civic skills is also a common practice in the literature (Verba, Schlozman, and Brady 1995). At the end of study 2, I also explained why this imperfect measure may attenuate the relationship between PM and civic participation. These test scores are norm-referenced and measure how students compare with their peers. (Ingels et al. 2007). As ELS:2002 did not collect data on 2004 reading tests, I used the 2004 ACT reading score as an intermediary outcome while controlling for school 2002 average reading scores from ELS:2002. As only students intending to attend college take

ACT or SAT, students with ACT scores may not be representative. In addition, 60% missing values for the ACT reading score will inevitably reduce the power of the tests. For this reason, I also presented the results using the reading score from ELS:2002. The limitation for using the reading score from ELS:2002 is that I cannot mitigate the concern of reverse causality even after controlling for the school average reading score in 2002.

The hours spent on and the categories of extracurricular activities in 2004 were used to capture the opportunities to develop civic skills in 2004. These measures took a well-established assumption that the more opportunities to hone civic skills a person is exposed to, the better the person's civic skills become (Verba, Schlozman, and Brady 1995). I removed sport-related extracurricular activities from the sum of extracurricular activities because they were shown to be less relevant for civic participation (McFarland and Thomas 2006). As parents' socioeconomic status can affect the development of students' interest and skills in civic participation (Feldman and Matjasko 2005; Musick and Wilson 2008), I adopted the common method of utilizing the school-centered socioeconomic status of student families to control for a student's relative capacity to participate in and benefit from civic participation relative to peers in the same school (Campbell 2009; Nie, Junn, and Stehlik-Barry 1996). Results remained similar in models using the absolute level of socioeconomic status.

Individual commitment to civic norms and school civic norms were used to capture the effects of values on civic participation. A strong commitment to civic norms will prompt one to engage in civic participation regardless of the results from a cost-benefit analysis (Dalton 2008). The degree of individual commitment to civic norms was measured through the first-factor score of individual

perceived importance of several behaviors ranging from helping others, supporting environmental causes to being an active and informed citizen, and being patriotic. Moreover, an individual may choose to participate to demonstrate compliance with civic norms and avoid social sanctions (Musick and Wilson 2008). To measure the effect of school civic norms independent from the effect of individual commitment to civic norms, I used an established method of utilizing the residual from regressing average individual commitment to civic norms at a school on the individual commitment to civic norms to reflect the school civic norms (Campbell 2009).

Controlling for Response Style

Failing to control for the response styles can reduce not only the efficiency of estimates but also their accuracy (see Appendix G for details). Response biases consist of two types. The first type refers to the tendency to systematically bias the response in one direction regardless of the item content. The second type is an inclination to adjust the responses according to content items to present a specific image about oneself (Watkins and Cheung 1995). The first type is called response style and the second response set (Lanyon and Goodstein 1982). To make the inference more accurate and efficient, the robust checks controlled for a series of measures of response styles (see Appendix G for how response styles may affect the estimations in the specific research setting). The controlled styles were the acquiescence response style (ARS), the tendency to agree with items regardless of content, disacquiescence response style (DARS), the propensity to disagree with items regardless of contents and noncontingent responding (NCR), the predisposition to respond to items carelessly and randomly and extreme response style (ERS), the inclination to endorse the most extreme response regardless of content (Baumgartner and Steenkamp 2001; see Appendix G for the questions used to construct the response styles).

Mid-point response (MPR) was not used here because respondents in an individualist country such as the U.S. were less concerned with expressing strong opinions and maintaining harmony and less likely to choose a middling response style (Hofstede 2001; Johnson et al. 2005).

Directly measuring response styles was recommended over the common factor technique because it specified the nature of the bias and would not confound the variations in methods with the variations in unspecified constructs (Podsakoff, MacKenzie, and Podsakoff 2012). In this case, running an exploratory factor analysis of all the student survey items used in the analysis yielded the first factor accounting for 34.7 % of the variance. As almost all the variables used in the study touch upon interpersonal interaction such as volunteering, extracurricular activities, and interacting with friends and parents. While the first factor may contain the common method bias, it may also reflect students' propensity for interpersonal interaction. I was interested in how performance management affects students' propensity for interpersonal interaction, which in turn, shapes students' volunteering behaviors. Specifically measuring the source of response biases was selected. In addition, controlling for multiple measures of response styles can enhance the stability of models and mitigate the common source bias, a threat to the internal validity of findings in public management research (Meier and O'Toole 2013; Podsakoff, MacKenzie, and Podsakoff 2012).

Methodology

Stata 16 was used to conduct path analysis to examine how school-level performance management in 2002 affected individual volunteering behaviors in 2004. Path analysis simultaneously analyzed the indirect effects of performance management on students' volunteering behaviors through math and reading abilities, civic skills, and individual commitment to civic norms. To account for the

multilevel nature of the dataset (schools and students), the estimation of standard errors was clustered at schools. I preferred clustered standard error over multilevel model for two reasons. First, the multilevel model in SEMs leads to difficulty in convergence for this study. Second, using clustering standard error instead of multilevel modeling also avoids adding complex and possibly inaccurate assumptions to the model (Primo, Jacobsmeier, and Milyo 2007).

The models also assumed that the type of extracurricular activities and the hours spent on extracurricular activities, two measures of the opportunities for developing civic skills, correlated. Stata's structural equation modeling (SEM) command was first used to assess the fit of the models. Due to the clustered estimation of standard errors, the only fit statistic available is Standardized Root Mean Square Residual (SRMR). As a reminder, I presented separate results from SEM using ACT reading scores and the ones from ELS:2002 as the intermediary measures of reading capabilities. The SRMR for the SEM with the ACT score and the one with the score from ELS:2002 was 0.038 and 0.039 respectively. Both were below the threshold of .08 and indicated adequate fit (Hu and Bentler 1999). As the multivariate normal distribution, a key assumption in the SEM command, was less likely to be met for the dependent variables, the command of GSEM was used in the final estimation. Logistic regression was used for the analysis of binary variable of volunteering; ordinal logistic regression was used for the analysis of ordinal variable of the frequency of community service; Possion regression was used for the analysis of the sum of volunteering categories because it was count data. Results from negative binomial regression for the sum of volunteering types were similar.

Quadratic terms of performance management components were included to account for the possible nonlinear relationships (Jacob 2005). All the equations controlled for common individual demographic variables, such as race and gender, and school demographics, such as the share of Black and Hispanic students and the share of students eligible for free or reduced-price lunch (Campbell 2008). To control for endogeneity that schools leading the implementation of performance management may differ from the rest, I included the school average of dependent variables in the base period (2002) in equations estimating test scores, participation in extracurricular activities, and commitment to civic norms.

In equations predicting volunteering behaviors, several control variables related to the network were utilized (see Table E-1 for details). Specifically, the model controlled for the opportunity of whether a school sponsored community service and the school's relationship with external stakeholders because both variables will affect students' opportunities for volunteering. Also included in the model were the strength of students' ties with friends, which has been shown to affect students' sense of solidarity and proclivity for volunteering (Musick and Wilson 2008). Moreover, the models accounted for the school standard deviation of students' commitment to the civic norms because the diversity of opinions is shown to affect participation (Campbell 2008; Mutz 2002). Students' college aspiration was included in the equation predicting volunteering behaviors and participating in extracurricular activities because the aspiration and the importance of volunteering in college admission may influence high school students' perceived benefit of engaging in these activities. Teachers' experience was added as a control in the equations estimating students' test scores for their importance in the estimation (V. E. Lee and Bryk 1989). The equation predicting the math test score contained the number of math-related courses taken by a student as a control (no corresponding variable exist for reading). Furthermore, I controlled for a student's perception of

welcome classroom climate and parents' participation in school activities (role modeling effects) in equations predicting students' commitment to civic norms because they have been shown to significantly shape students' proclivity for civic participation (Campbell 2009; Musick and Wilson 2008). Control variable related details can be founded in Appendix F.

Four robustness checks were conducted to verify the internal validity of findings. As both extracurricular activities and volunteering can develop civic skills and foster disposition toward civic participation (Musick and Wilson 2008), it may be hard to gauge the direction of the relationship between participation in extracurricular activities and volunteering in the same period. In addition, even after controlling for a comprehensive set of variables, there can still be a lingering concern about the selection effect of the fundamental difference between students participating in extracurricular activities and volunteering and those who are less active (McFarland and Thomas 2006). To address these issues, I added student volunteering behaviors in the base period (2002) and replaced volunteering measures in 2004 with the same ones in 2006. Controlling for volunteering behaviors in the base period can mitigate the concern of selection bias. The temporal sequence in time between participation in extracurricular activities in 2004 and volunteering behaviors in 2006 reduces the concern of reverse causality. The second robustness check used state dummies to control for the between state variation in the requirement of community service (Shapiro 2018) and verified whether the effects of predictors for students' volunteering behaviors still hold. The third robust test took into account aforementioned response styles. The fourth robust test distinguished between within-school variation and between school variation and estimated the indirect effect of school-centered measures and school averaged measures of math scores, frequency, types of extracurricular activities, and endorsement of civic norms on volunteering behaviors (Zhang, Zyphur, and Preacher 2009).

Results

To test the significance of the indirect effects of performance management on volunteering behaviors, I relied on both the distribution of product method and test of joint significance method. The distribution of product method has been shown to generate the confidence interval with more accurate Type 1 error rate and power (MacKinnon et al. 2007). The test of joint significance is common and is shown to perform relatively well (Hayes and Scharkow 2013). In applying the test of joint significance, I used five percent significance level at two-sided tests as the threshold, except for the equations involving ACT reading scores. Given the low power of tests involving ACT reading scores, I lowered the threshold to ten percent. Using the R package of RMediation for the distribution of product method (Tofighi and MacKinnon 2011), I produced the 95% confidence interval for the indirect effects of performance management on the three measures of volunteering (results can be obtained upon request).

The results (Table 6 and 7) confirmed Hypotheses 1 that performance management enhances civic participation by improving students' math and reading abilities. The linear term of user-oriented component of performance management significantly and consistently increased student probability of volunteering, frequency of volunteering, and total types of volunteering (p values for the average marginal effect of math score on all three volunteering behaviors <0.01) by improving students' math score (p-value: 0.022). In SEM, using ACT reading score as the intermediary variable, the quadratic term of user-oriented component was also positively associated with the probability of volunteering (p-value for the marginal effect of ACT reading score on volunteering: 0.055) through its positive effect on ACT reading score (p-value for the effect of UOC^2 on ACT reading: 0.069).

In SEM using reading scores from ELS:2002 as the intermediary variable, the linear term of the user-oriented component was positively associated with volunteering probability and the total sum of volunteering types (the P-value for the effect of reading scores on volunteering probability <0.01 and P-value for the effect of reading scores on volunteering types:0.039) through its effects on reading scores (p-value: 0.048). Moreover, the quadratic term of the provider-oriented component was positively associated with volunteering probability and volunteering types through its positive effect on reading scores (p-value: 0.020). The qualitative results from the distribution of the production method remained similar. The magnitude of all indirect effects of performance management on volunteering behaviors through its effects on cognitive abilities was less than 0.09 standard deviations, the threshold for the medium indirect effect (Kenny 2018), and, therefore, small.

The results indicated mixed evidence for Hypothesis 2 about the negative indirect effect of performance management on volunteering through extracurricular activities. The linear term of provider-oriented component was consistently and negatively associated with volunteering behaviors (the p-values for the effect of frequency of extracurricular activities on all three types of volunteering behaviors<=0.01) through its negative effects on the frequency of extracurricular activities (p-value: 0.03). The magnitude of all the negative indirect effects through participation in extracurricular activities was smaller than 0.09 standard deviations and, therefore, small (Kenny 2018). Contrary to Hypothesis 2, the user-oriented component of performance management did not negatively correlate with volunteering behaviors at a five percent significance level. The results did not support Hypothesis 3. At the five significance levels, performance management did not correlate with students' commitment to civic norms and, therefore,

did not have a significant indirect effect on this volunteering through this pathway. The results from the distribution of the product method were consistent.

In summary, the findings indicate a statistically significant, consistent, but small positive indirect effect of the user-oriented component of performance management on volunteering behaviors through both reading and math test scores. Some evidence also exists for the positive indirect effect of the provider-oriented component of performance management on volunteering behaviors through reading scores when the power of the test was high. Moreover, results also indicated a statistically significant, consistent, but small negative indirect effect of the provider-oriented component on volunteering behaviors through participation in extracurricular activities. The user-oriented component did not negatively affect volunteering behaviors. Neither the user-oriented nor the provider-oriented components significantly affected students' commitment to civic norms. The results from robust checks remained similar and proved the persistent effect of performance management on students' volunteering behaviors (see Table H3-13 in Appendix H). The consistency of results in models controlling for response style also is also found in OECD's reviews on studies on people's institutional and interpersonal trusts (OECD 2017).

Table 6. The Abridged Results with ACT Reading Score at the Intermediary Variable

VARIABLES	ACT04	Math04	Extratype04	Extrahour04	Civic04	Volr04	Voltype04	Volfreq04
User-oriented component (UOC)	-0.0982	0.323**	-0.0322	0.00177	-0.00718			
	(0.148)	(0.141)	(0.0408)	(0.0321)	(0.0126)			
UOC^2	0.190*	-0.0966	0.00293	-0.0184	0.0101			
	(0.104)	(0.0934)	(0.0287)	(0.0206)	(0.00891)			
Provider-oriented component (POC)	-0.0620	0.0431	-0.0380	-0.0863***	0.0178*			
	(0.163)	(0.126)	(0.0342)	(0.0295)	(0.0107)			
POC^2	-0.0467	0.129	-0.0377*	-0.0292	0.00334			
	(0.101)	(0.0801)	(0.0217)	(0.0179)	(0.00710)			
Math score from ELS:2002						0.0383***	0.0134***	0.0279***
						(0.00817)	(0.00308)	(0.00642)
ACT reading score						0.0238*	0.00316	-0.0117
						(0.0124)	(0.00412)	(0.00885)
School-centered individual socioeconomic score	1.791***	3.406***	0.517***	0.407***	0.0314***	0.257***	0.0540*	0.0853
	(0.185)	(0.179)	(0.0389)	(0.0348)	(0.0122)	(0.0856)	(0.0303)	(0.0710)
Individual Sum of extracurricular activity types						0.334***	0.111***	0.215***
						(0.0380)	(0.00795)	(0.0218)
Individual Frequency of engaging in extracurricular activities						0.126***	0.0637***	0.0785***
						(0.0261)	(0.0101)	(0.0236)
Individual commitment to civic norms						0.712***	0.308***	0.957***
						(0.103)	(0.0424)	(0.0835)
Number of observations used in the specific equation	2740	6950	8080	8060	5980	2450	2460	2440
Number of observations ever used in the whole model				8,260				

Robust standard errors in parentheses (two-sided tests)

*** p<0.01, ** p<0.05, * p<0.1

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics.
(2) Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2000-2001, 2000-2001, Version 1, U.S. Department of Education, National Center for Education Statistics. The next nine tables use the same dataset.
Note. All models had the following variables as controls: race, gender, and school-centered socioeconomic status, school share of African American and Hispanic students, and share of students eligible for free or reduced-price lunch. The equations estimating test scores also accounted for teachers' experience, and school average test scores in the base period (2002). The equation predicting the math score also included the number of math-related courses (a corresponding variable related to reading does not exist).The equations estimating participation in extracurricular activities also controlled for the strength of friends' ties, college aspiration, and school average participation in extracurricular activities in the base period (2002). The equation estimating students' commitment to civic norms also controlled for parents' role modeling effect, a student's perception of welcome class climate, and school average perceived importance of helping others in the base period. The equations estimating volunteering behaviors took into account students' college aspiration, school civic norms, the standard deviation of commitment to civic norms at schools, school relationship with external stakeholders, and whether the school sponsored community service programs.
To comply with the requirements of using restricted dataset of ELS:2002, all the unweighted sample sizes were rounded to the nearest 10. Full results were presented in table H-1 in Appendix H.

Table 7. The Abridged Results with ELS Reading Score at the Intermediary Variable

VARIABLES	Read02	Math04	extratype04k	extrahour04	civic	volunteer04	voltype04	volfreq04
User-oriented component (UOC)	0.171**	0.323**	-0.0322	0.00177	-0.00718			
	(0.0861)	(0.141)	(0.0408)	(0.0321)	(0.0126)			
UOC^2	-0.0520	-0.0966	0.00293	-0.0184	0.0101			
	(0.0556)	(0.0934)	(0.0287)	(0.0206)	(0.00891)			
Provider-oriented component (POC)	-0.000142	0.0431	-0.0380	-0.0863***	0.0178*			
	(0.0759)	(0.126)	(0.0342)	(0.0295)	(0.0107)			
POC^2	0.112**	0.129	-0.0377*	-0.0292	0.00334			
	(0.0478)	(0.0801)	(0.0217)	(0.0179)	(0.00710)			
Math score from ELS:2002						0.0333***	0.0132***	0.0177***
						(0.00456)	(0.00229)	(0.00408)
Reading score from ELS: 2002						0.0207***	0.00454**	-0.00101
						(0.00485)	(0.00220)	(0.00402)
School-centered individual socioeconomic score	3.933***	3.406***	0.517***	0.407***	0.0314***	0.242***	0.0795***	0.0936*
	(0.181)	(0.179)	(0.0389)	(0.0348)	(0.0122)	(0.0553)	(0.0222)	(0.0478)
Individual Sum of extracurricular activity types						0.324***	0.112***	0.221***
						(0.0225)	(0.00722)	(0.0141)
Individual Frequency of engaging in extracurricular activities						0.125***	0.0713***	0.0965***
						(0.0174)	(0.00807)	(0.0156)
Individual commitment to civic norms						0.433***	0.282***	0.828***
						(0.0587)	(0.0289)	(0.0542)
Number of observations used in the specific equation	6740	6950	8080	8060	5980	5960	6020	5960
Number of observations ever used in the whole model				8270				

Robust standard errors in parentheses (two-sided tests)

*** p<0.01, ** p<0.05, * p<0.1

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics.
(2) Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2000-2001, 2000-2001, Version 1, U.S. Department of Education, National Center for Education Statistics. The next nine tables use the same dataset.
Note. All models had the following variables as controls: race, gender, and school-centered socioeconomic status, school share of African American and Hispanic students, and share of students eligible for free or reduced-price lunch. The equations estimating test scores also accounted for teachers' experience, and school average test scores in the base period (2002). The equation predicting the math score also included the number of math-related courses (a corresponding variable related to reading does not exist).The equations estimating participation in extracurricular activities also controlled for the strength of friends' ties, college aspiration, and school average participation in extracurricular activities in the base period (2002). The equation estimating students' commitment to civic norms also controlled for parents' role modeling effect, a student's perception of welcome class climate, and school average perceived importance of helping others in the base period. The equations estimating volunteering behaviors took into account students' college aspiration, school civic norms, the standard deviation of commitment to civic norms at schools, school relationship with external stakeholders, and whether the school sponsored community service programs. To comply with the requirements of using restricted dataset of ELS:2002, all the unweighted sample sizes were rounded to the nearest 10. Full results were presented in table H-2 in Appendix H.
To comply with the requirements of using restricted dataset of ELS:2002, all the unweighted sample sizes were rounded to the nearest 10. Full results were presented in table H-1 in Appendix H.

Discussion

The findings show the value of applying policy feedback theory in public management and echo the call for studying how variation in policy implementation shapes citizenship (Moynihan and Soss 2014). Motivated by the lack of empirical research on the impact of management on citizenship, I draw on policy feedback theory to hypothesize mechanisms through which performance management focusing on test scores can affect civic participation. The results show mixed results and a more balanced assessment of the relationship between performance management and civic participation. On the one hand, performance management was shown to enhance civic participation by improving students' cognitive abilities and better preparing them to perform various tasks involved in participation. On the other hand, performance management was shown to weaken civic participation by reducing students' participation in extracurricular activities where they practice civic skills. The findings imply that performance management emphasizing workforce preparation both complements and competes with the school mission of building democratic citizenship. Like public education, the implementation of federal policy/programs can be highly heterogeneous (Herd and Moynihan 2018). Hence, extending policy feedback theory to management in other policy areas may also help scholars build a more comprehensive and balanced assessment of the impact of policy and its management on citizenship.

In applying policy feedback theory to the study of public management, taking a coproduction perspective can yield insight. In this study, I organize performance management along the axis of a service user (students) versus service provider (teachers and principals). This coproduction perspective helps us account for heterogeneous effects of performance management on measured outcomes like cognitive abilities and unmeasured outcomes like civic

skills. First, the user-oriented component was shown to have a more consistently positive effect on cognitive abilities than the provider-oriented component because this component may provide more informative performance feedback documenting the true variation in performance. As a reminder, the frequency of testing is the main informational subcomponent for the user-oriented component while teacher-evaluation is the main one for the provider-oriented component. Taking tests often yields information students and teachers can utilize to improve results (Andersen and Nielsen 2019), not least because test performance can reveal meaningful variation in how students and their teachers fare (Rockoff et al. 2012). By contrast, teacher evaluation often fails to distinguish among poor teachers, average teachers, and outstanding teachers because most teachers are rated as either proficient or exemplary (Weisberg et al. 2009). Due to a lack of variation, a teacher does not know how he/she is doing in comparison with peers from teacher evaluation, not to mention how to improve their performance.

Taking the user versus provider division on performance management also assists in searching for the explanations for mixed effects of performance management on unmeasured outcomes like participation in extracurricular activities and the development of civic skills. I found that the provider-oriented component of performance management had a significant negative association with participation in extracurricular activities, but the user-oriented component had an insignificant impact. A possible reason is that the users (students) are subject to another set of a performance management system that counteracted the omission of unmeasured outcomes caused by the studied performance management system (Kelman and Friedman 2009). Alongside the studied performance management in the paper which focused on state educational targets for students, schools usually have another performance management system focusing on student college readiness. The performance management system for college

admission assigns equal, if not more, weight to extracurricular activities than to students' academic performance. If I make the reasonable assumption that the degree of implementing the two performance management systems covaries, then s may expect the effects of these two performance management systems to cancel each other out, thus indicating an aggregated null effect. By contrast, public school employees are not subject to another performance management system preventing the omission.

Taking a co-production approach can also help us put either the small or null effect of performance management on civic participation in perspective because other drivers can be more impactful for educational outcomes. Whether it is measured outcomes like reading or math test scores or unmeasured ones like extracurricular activities and development of civic skills, achieving desired results requires close collaboration between students, the service users, and public employees at schools, service providers (S. P. Osborne, Radnor, and Strokosch 2016). A consistent feature of public education as a coproduction is that students' demographic characteristics are more impactful factors for educational outcomes than what measures public employees have taken at schools (Moe 2009). The results confirm this feature and show that the association of student and school demographic variables with test scores and participation in extracurricular activities were often much larger than the associations of performance management (see columns 1-4 in Table C-1 and Table C-2). The null effect of performance management on students' commitment to civic norms also confirmed complex factors influencing people's attitudes toward participation (Soss and Schram 2007). The significant effect of students' perception of class climate and parents' role modeling echo previous findings that students' interaction with classmates, teachers, and parents are more proximate causes for students' proclivity for civic participation (Campbell 2009; Musick and Wilson 2008).

There are limitations to the current study and room for future studies. First, using participation in extracurricular activities as a measure of civic skills may not be reliable. The proxy measure depends on the assumption that participating in organized activities will help people exercise and further develop their civic skills (Verba, Schlozman, and Brady 1995). The degree to which participating in organized activities contributes to the development of civic skills may vary across people and constitutes a source of classical measurement error. The classical measurement error will attenuate the relationship and may explain the small effect size. To address the limit, I added the ACT writing scores as an intermediary outcome measure of civic skills and found insignificant effects of performance management on students' ACT writing scores possibly due to the low power of the tests. Given the narrow focus of standardized tests, a better measure may be students' grades in English courses (Condon 2015). However, using students' grades on English coursework may raise the question of whether these grades can be compared across schools. In addition, ELS: 2002 only has the data on cumulative GPA rather than GPA on a specific course. Future research can explore whether more reliable measures of civic skills change the magnitude of the indirect negative effect of performance management on volunteering behaviors.

Second, the study does not rely on exogenous change and thus cannot make strong causal inferences. I went to great lengths to enhance internal validity by conducting a series of robust checks (see method section). However, it is still possible that some unobserved variables accounted for the significant relationships between performance management and civic participation. Given the importance of this topic, this article only marks the beginning of empirical studies on the impact of performance management on civic participation. Future studies can try to take advantage of exogenous changes in the performance measurement system

to see how the variation affects civic participation. One promising source of exogenous changes is the enhancement of teacher evaluation and linking evaluation results to decision making after Obama administration incentivized states to adopt these measures to increase their chances to win Race to the Top grants.

Conclusion

As performance management expands to almost all aspects of modern governance, the tension between the "simplicity" of performance management and the "complexities" of modern governance has become salient (Moynihan et al. 2011). A key attraction of performance management is that it provides a tool to replace multiple, vague, and often contradictory goals with a few, clear, and consistent ones (Moynihan 2008). The complexities of modern governance such as multiple principals and the need for honoring democratic values often belie the central appeal of performance management (Moynihan et al. 2011).

A concrete example of the criticism above is the purported negative effects of performance management on students' civic participation in public education. The findings of this article support the concern that overemphasizing performance measures based on test scores can detract from the mission of public education to develop democratic citizenship (R. Rothstein, Jacobsen, and Wilder 2008). A decline in civic participation can undermine the fundamental goal of performance management—restoring public trust through better government performance (Snyder, Saultz, and Jacobsen 2017; Wichowsky and Moynihan 2008). However, it is important to put the negative effects in perspective. Performance management does enhance civic participation by improving students' cognitive abilities to process information related to participation. The mixed-effects suggest that performance management focusing on building a

skilled workforce can both complement and conflict with the mission of public schools to foster democratic citizenship.

SUMMARY AND DISCUSSION

Overall, this dissertation is motivated by the possibility of conflict between performance management and democratic values. The findings of the dissertation show that the concerns of the conflicts are warranted. Due to the incompleteness of performance measures in capturing all the valuable goals and measuring all citizens' wellbeing (Dixit 2002; Holmstrom and Milgrom 1991), performance management can lead to prioritizing some high-value citizens at the expense of others and improvement in the measured outcomes at the expense of unmeasured but important outcomes important to democracy such as civic participation. Still, the dissertation shows that user-oriented component of performance management helps struggling citizens and mitigates the iniquity of prioritizing high-value citizens. In addition, the user-oriented component promotes civic participation by equipping students with the cognitive skills needed for participating in public affairs. These findings imply that blanket statements about the effects of performance management are neither accurate nor helpful. Performance management consists of different components and the associations of these components with outcomes can be heterogeneous. In this study, I found that the user-oriented component of performance management generates more positive outcomes for equity and civic participation.

Complement Policy Reforms With Managerial Reforms to Advance Democratic Values

These findings also support the impacts of management on persistent challenges beyond policy. The recent protests for racial justice again highlight the persistent racial educational achievement gaps between whites and minority groups in areas like Minneapolis (Grunewald and Nath 2019). With few exceptions focusing on the management of policy such as pairing minority

students with minority teachers (Gershenson, Hansen, and Lindsay 2021), the search for solutions revolves around policies (Fryer 2017). The selected solutions out of the search tend to be incomplete and insufficient for shrinking the achievement gap and promoting equity for two reasons. First, as is demonstrated in this dissertation, significant variations exist between states in implementing federal policies (Education Week 2003). The differences in managerial practices at the organizational level are even larger and can account for the outcomes of service delivery, especially for disadvantaged groups (Heinrich et al. 2014). Hence, leaving managerial practices out of the search limit policymakers' ability to tackle the inequity.

Second, policy solutions to flaws in previous policies are always susceptible to new problems. For instance, the Every Student Succeeds Act eliminates the use of adequate proficiency progress partly in response to various performance perversities like prioritizing high-value citizens. To encourage public schools to take care of every student, especially those struggling ones, many districts have adopted value-added models and used teachers' value-added as the new primary criteria for holding schools accountable. By controlling common confounders for students' achievement such as demographic variables and classroom characteristics, value-added model has the promise of isolating teachers' value-added to students' learning, thereby removing the incentive to prioritize students close to any arbitrary cut-off points and mitigating the inequity. However, it is still debatable whether the model accurately captures teachers' contribution to students' academic performance (Chetty, Friedman, and Rockoff 2014; J. Rothstein 2016). The uncertainty of value-added modeling also raises another equity question of whether schools and teachers got punished for serving struggling students who take more time to achieve the same degree of performance improvement. Because of the imperfect nature of performance measurement, public employees may adopt new gaming strategies to cope with the

accountability pressure (Heinrich and Marschke 2010). In this case, public employees may allocate time to students with the highest growth potential away from others.

Incorporating managerial solutions can help break the vicious cycle of detecting gaming in previous policies and fixing the problems in the next round of legislation only to discover public officials taking advantage of the loopholes again in the new system. Public employees want to assist citizens needing help the most (Jilke and Tummers 2018). They choose to game the system primarily as a way to cope with accountability pressure for goals they cannot attain (Terman and Yang 2016). Hence, enabling public managers to attain results can tackle the root cause of gaming. To this end, the dissertation suggests that providing more accurate performance feedback and fostering a more supportive environment for performance management can contribute to performance improvement and mitigate the overlooking of struggling citizens caused by incomplete performance measures. Other studies also found that granting managers autonomy in managing public organizations can improve organizational performance and mitigate cream skimming and goal displacement in performance management (Nielsen 2014b; Han and Wang 2021). Together, these findings suggest that when public employees' intrinsic motivation for competence and autonomy are satisfied by performance management practices, their public service motivation will be enhanced and they will strive to serve every citizen despite the shortcomings in policy designs.

Practitioners and researchers in public education also realize the importance of complementing policy reforms with managerial reforms. The improvement science, a disciplined study of how to get better in managing school is getting traction in public education (Bryk et al. 2015; Bryk 2020). As Figure 4 shows, improvement science consists of six principles.

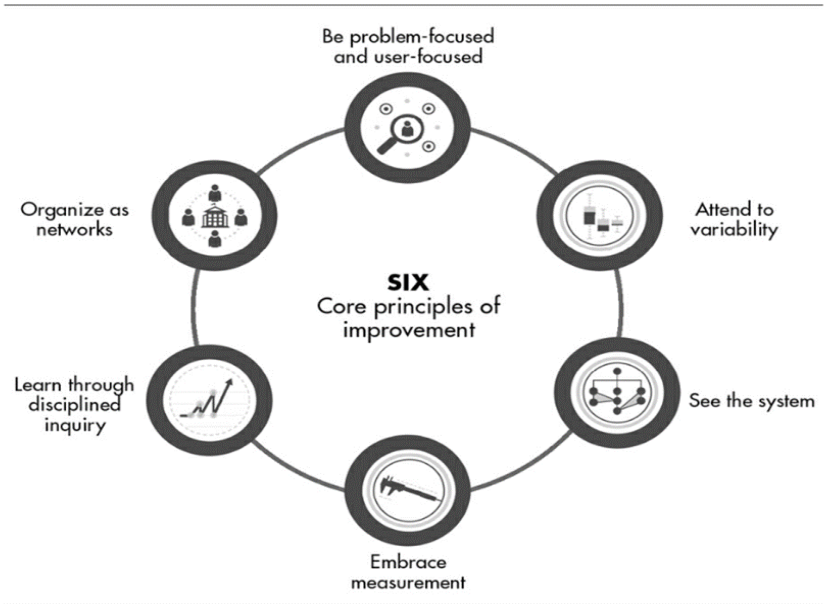

Figure 4. Principles of Improvement Science

Source: The figure is taken from chapter one of the book Improvement In Action (Bryk et al. 2015).

Note. The Kindle version of the book does not show the page of the figure.

Focusing on problems and users suggests avoiding jumping into solutions without analyzing and understanding the actual problems. Attending to variability means searching for performance outliers and studying reasons for the deviation. Seeing the system refer to understanding fully the underlying process, norms, and organizational arrangements that produce unacceptable outcomes. Embracing measurement and learning through disciplined inquiry is about generating disaggregated performance information and participating in an iterative cycle of developing, testing, revisiting, and possibly revising changes based on data. Organizing as a network requires building a problem-solving network that shares similar goals and collaborates to develop and promote knowledge and practices for improving performance. The improvement science focuses on how to improve school management as much as on designing new programs (Bryk et al. 2015; Bryk 2020). The emphasis on data as the basis for decision-making and

continuous improvement shows that the improvement science incorporates the principle of performance management in its pursuit of improving school management.

Take a Systematic Approach to Performance Management

The findings suggest that future studies on the relationship between performance management and democratic values may benefit from a systemic approach. The systemic approach includes at least fourth parts. The first part is to adopt the common practice in strategic human resource management research and incorporate bundles of practices related to performance management in the model (Boselie, Van Harten, and Veld 2021). The principal component analysis in the dissertation indeed supports assumptions for use of the bundle approach: a variety of practices are interwoven to exert influence. The analysis supports goal setting, performance monitoring, attaching consequences to performance, and managerial empowerments to be highly connected and play indispensable roles in unlocking the potential of performance management. While an isolated study may be strong in making causal inferences, it cannot adequately account for the synergy between a bundle of practices related to performance management.

Second, bundles of performance management practices need to incorporate the coproduction nature of public service and cover important stakeholders. As public education is a typical example of co-production, the dissertation measures goal setting, performance monitoring, and attaching consequences to performance for both students (service users) and teachers and principals (service providers). Without the comprehensive measurement of performance management, I could not discover the nuanced relationships between performance management and equity. Neither can I adequately account for the mixed effects of performance management on civic participation.

Given the prevalent requirement of co-production in public services (S. P. Osborne, Radnor, and Strokosch 2016), measuring all the tenets of performance management for both service providers and service users in other fields will yield more nuanced findings beyond public education. Stakeholders with different perspectives on the drivers of performance increasingly work together to deliver public service (Amirkhanyan, Kim, and Lambright 2014; Ansell and Gash 2008; Hamilton et al. 2007). The collaborative nature of work and diverse perspectives of stakeholders means that focusing on performance management for one stakeholder is likely to have an incomplete understanding of the drivers for organizational performance.

Third, the findings also indicate that studies of performance management need to do more to explore whole cycles of changes from the design of performance management to the alternations in behaviors and processes, and the attainment of desired outcomes. The dissertation found that lack of attention to civic participation in the design of high-stakes performance management at American high schools did reduce students' civic participation by hindering students' development of civic skills, it also enhanced students' civic participation by improving their cognitive abilities. The mixed effect of performance management calls attention to the importance of studying the assumed relationships among design, processes, and outcomes. If the measured process and outcomes complement the unmeasured ones, the omission of certain outcomes at the design of performance management may not lead to net negative associations of performance management with the unmeasured outcomes (Kelman and Friedman 2009). The importance of studying the whole process can also be seen in the finding that performance management spawned changes in processes aiming to help disadvantaged groups, but the changes seem to be ineffective in achieving the outcomes of improving disadvantaged groups'

performance (Heinrich et al. 2014). Other studies on public policies echo the importance of understanding how policy inputs can translate to desired outcomes (Bryk et al. 2015). Given the variation in implementing the policy, it is not surprising that the outcomes of implementing the same policy differ. The variation in implementation and outcomes provides a fertile ground for scholars of public management to study pathways from input and outcomes and identify conditions conducive to the successful translation of input to desired outcomes.

Examining the mechanism of how policy implementation affects policy outcomes can also serve as an alternative method of enhancing the external validity of research findings. The recent popularity of experimental methods in public management studies has spawned discussion on the external validity of experimental findings. A common boilerplate response to questions of external validity is that more replications of similar studies should be needed in the future. However, the constraints on resources also make the replication infeasible. An alternative route to enhance the external validity of research findings is to examine the mechanisms of translating input to outcomes and see whether the key mechanism can still be at play in other places (Ludwig, Kling, and Mullainathan 2017). Testing the key mechanism may require fewer resources than full replications of experiments. For instance, the mechanism from providing accurate performance feedback on public employees to better organizational performance is that (1) public employees will adjust self-assessment according to performance feedback (Andersen and Nielsen 2019) (2) public employees will find and implement effective solutions to the shortcomings revealed in performance feedback (Holm 2018). Understanding whether these two mechanisms are at play can help policymakers and researchers evaluate whether the intervention of providing accurate performance feedback can lead to better performance.

The fourth part of a systemic approach is to understand the connection between the past and present. As is demonstrated in the dissertation, performance management in public education has been implemented for at least three decades now. Many measures of performance management have been tried before and past experience can inform the present policy design. For instance, the emphases on disaggregated performance information and accountability to performance in No Child Left Behind continue in the Every Student Succeeds Act. The continuation of past policies to the present suggests that adding more measures does not necessarily generate better results. One recommendation of the dissertation is that performance management should aim to provide feedback telling public employees how they are doing and what they need to do next. Policymakers may adopt policy recommendations and generate a new system to provide performance feedback. Jumping into creating new programs without understanding past policies is likely to be ineffective in attaining the desired outcomes.

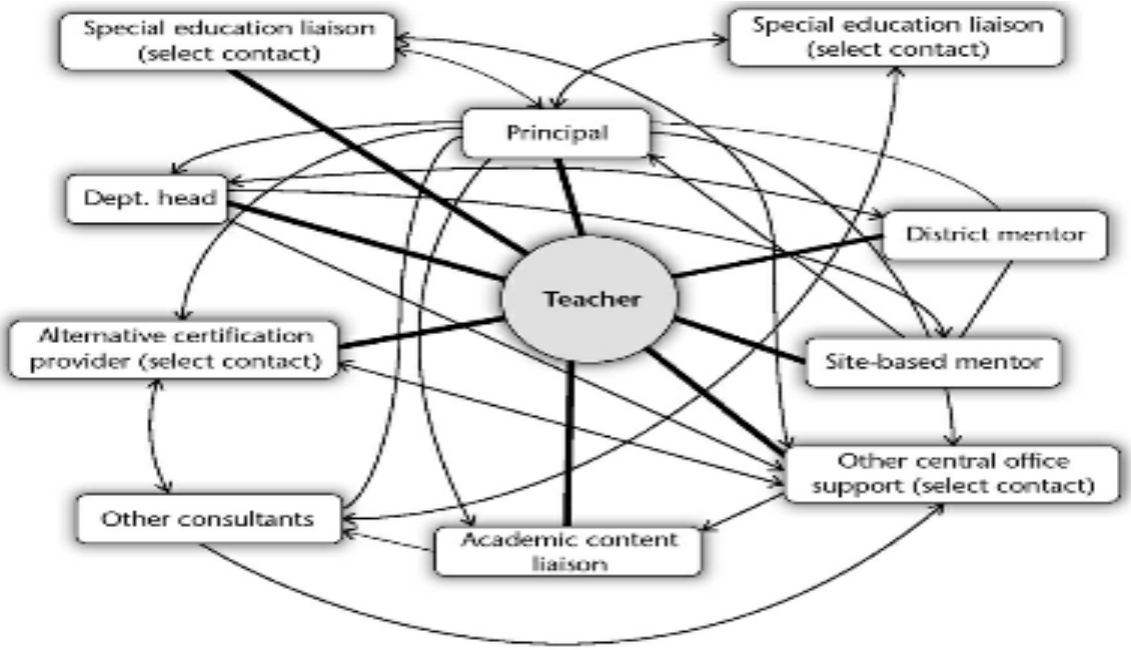

Figure 5. The Blizaard of Feedback for Some Teachers in Baltimore City

Source: The figured is taken from page 30 of the Book *Learning to Improve* (Bryk et al. 2015)

For instance, as is illustrated in Figure 5, some teachers in Baltimore public schools face a maze of feedback. Creating a new system of feedback giving will make feedback received by teachers confusing and possibly overwhelming. A more sensible approach is to comb through the different feedback and make the feedback coherent and manageable to teachers (Bryk et al. 2015). The lesson is that the solution to the existing system may lie in fixing the existing system itself rather than generating a new system.

Still, I also acknowledge that this lesson may not apply in many other settings where the past leaves few systems from which to start. For instance, in the same city of Baltimore, some public teachers receive no feedback at all (Bryk et al. 2015). Hence, there is not an existing system to be improved at some public schools. The various settings call for managerial empowerment, a key component of performance management. With a clear focus on and accountability to performance, empowered public managers will tailor programs to their clients' needs and adapt the programs to diverse and shifting environments (Ouchi and Segal 2003).

Advance Contingency-Based Understanding of Performance Management

The discussion above also raises the question of what others can learn from this dissertation about the relationships between performance management, gaming behaviors, and democratic values in other settings? To answer this question, I start with two observations on performance management in both developing and developed countries. First, public employees are more likely to resort to gaming behaviors when they do not identify with performance goals and have limited capacity to attain multiple, competing, and high-stakes goals (Soss, Fording, and Schram 2011; Terman and Yang 2016; Gao 2015). Second, fixing the institutional loopholes that make existing gaming possible without addressing the other aforementioned problems will make gaming behaviors evolve rather than disappear (Heinrich and Marschke 2010; Gao 2020).

Hence, the question for how to mitigate gaming in performance management becomes under what conditions will public employees be more likely to identify with performance goals and perceive performance management as a tool to attain challenging but realistic goals?

To answer these questions, I draw on the human resource management (HRM) bundles theory (Boselie, Van Harten, and Veld 2021; Huselid 1995) and public management context matrix developed by O'Toole and Meier (2015). The theory of HRM bundles holds that HRM practices will be satisfying to employees and improve organizational performance if practices are internally coherent (horizontal fit), aligned with organizational strategy (vertical fit), and compatible with the organizational system including resources, structure, and culture (organizational fit) and external environment (environment fit). Studies show that achieving internal and external fit for performance management can improve employees' satisfaction with performance management and enhance employees' organizational commitment (Decramer, Smolders, and Vanderstraeten 2013; Van Waeyenberg et al. 2017). The hypotheses on how contexts shape managerial impact suggest that attaining organizational and environmental fits can increase the marginal return of managerial activities (O'Toole and Meier 2015). When organizations attain the four fits, performance management will likely deliver performance improvement and boost public employees' morale, reducing the motivation for gaming the system to cope with performance management.

Performance management will likely have buy-in from public employees and improve organizational performance when the practices of performance management are in line with the doctrine of performance management. The doctrine of performance management holds that managers need to be empowered to manage for results in exchange for accountability to results (Moynihan 2008). In reality, accountability to results is often implemented without granting

managers the autonomy to attain results (Jakobsen and Mortensen 2016). Under such circumstances, gaming may become more rather than less likely because of the lack of alternative means to achieve goals. Results from the principal component analysis in the dissertation also suggest that managerial empowerment and attaching consequences to results for teachers go hand in hand in the service-provider component, further validating the coherence of performance management doctrine. In another paper, I find that fully implementing the doctrine of performance management improves average student performance in low-stake subjects, and student subgroup performance not emphasized in performance assessment, thereby mitigating goal displacement (Han and Wang 2021). Moreover, learning from performance feedback is another key tenet of performance management doctrine (Moynihan and Landuyt 2009). Overall, these dissertation findings suggest that effective learning from timely and accurate performance information that can distinguish between outstanding and mediocre public employees is necessary for fostering learning and countering the incentive to prioritize high-value citizens. Empowering managers can also increase public managers' organizational commitment (Osborne 2017) and encourage them to accumulate expertise to interpret and use performance information (Andersen and Moynihan 2016).

Implementing performance doctrine is more likely to succeed if sufficient resources and support are allocated to assist the implementation. Implementing tenets of performance management can bring about two risks. First, the complexity and volume of transformation may overwhelm public employees and cause worse performance (Hemphill et al. 2010), incentivizing coping with performance pressure through gaming. Second, unethical public employees may use the newly granted autonomy to game the system and advance self-interest (Schuster, Meyer-Sahling, and Mikkelsen 2020). To mitigate the risks, competent and ethical public employees

need to be in charge of implementing performance management. In addition, organizations must provide sufficient resources to build the capacity for transitioning to manage for results and keep the workload for public employees at a reasonable level (Han and Wang 2021). Moreover, the dissertation also suggests that a supportive environment is fostered when organizations support public employees to improve performance in cases of missing their performance targets and recognizing their achievement in cases of exceeding their performance targets. Supportive organizational environments can contribute to changing public employees' perception of performance management from a damper to an enabler for their internal need for autonomy, competency, and relatedness (Corduneanu, Dudau, and Kominis 2020). When public employees feel that performance management helps them satisfy their basics needs, their public service motivation is enhanced. With strong public service motivation, public employees tend to go the extra mile to serve the public even though it may not align with their self-interest (Moynihan 2010), thereby mitigating the possibility of gaming.

To continually improve organizational performance, and reduce the need for gaming, organizational strategies and an emphasis on performance management practices need to be aligned with the stage of performance management. Strategies for public organizations can be grouped into four types: (1) prospecting: always seeking new ways of providing services, running organizations, and interacting with other organizations (2) defending: taking a conservative view of changes and focus on improving the efficiency of existing operation (3) analyzing: taking an in-between approach (4) reacting: seldom adjusting until forced to do so by environmental pressure (Boyne and Walker 2004). Due to the constraints on public organizations, few will fully embrace prospecting as the primary strategy. In the delivery of core public organization services, defending was shown to be the best organizational strategy because

the stability and continuity enables public organizations to focus on improving their core business (Meier et al. 2006). At the start of performance management, most organizations likely take an analyzing approach. Although they are still bound by laws and public expectations to provide service and goods to all eligible citizens, they will use strategic planning to clarify missions and explore ways of fulfilling the missions. Once the strategic plan is formulated and connected to performance goals for organizational operation, organizations may benefit from the stability and a relentless focus on improving the efficiency of attaining performance goals brought by defending strategy (Behn 2014). The dissertation suggests that process changes spawned by performance management were not effective for generating the outputs such as improving traditionally disadvantaged students' test scores, public organizations must shift back to analyzing strategy to evaluate the assumptions underlying the logic model from the process changes to output and explore alternative ways of producing the outputs. Alternatively, others argue that public organizations also need to reutilize the analyzing strategy when meeting the output targets not improve the outcomes (Kleine 2018). While all performance management practices may still at work, deploying the analyzing strategy prioritizes revisiting the assumptions in the strategic plans, using performance evaluation to which part of the logic models are mistaken, and exploring alternative ways of providing services and goods if necessary (Han and Moynihan 2021).

As for the environmental fit, few studies have empirically examined the relationship between how variation in political and institutional environments shape outcomes of performance management, not to mention specific gaming behaviors. One reason for the paucity of research is that the adoption of performance management is limited in countries with a strong Weberian tradition (Pollitt and Bouckaert 2017). Performance management is often considered

part of the New Public Management Movement that holds that practices in the private sector can be used to improve government operation. This claim was appealing to many politicians in countries such as United Kingdom, New Zealand, Canada, and United States where the general public perceives blurred lines between private and public organizations and laws are usually in the background than in the foreground. By contrast, this view is met with skepticism in Weberian countries such as France, Germany, and Nordic countries where states play a unique and central role in governance, and rule-following and hierarchy are dominant ways of governance (Pollitt and Bouckaert 2017). In addition, many countries in continental Europe have a corporatist tradition of governance through consultation and consensus and some agencies balk at the prospect of divisiveness and confrontations brought by performance management. For all these reasons, performance management is more fully implemented in countries embracing New Public Management reforms and most of the studies on gaming in performance management in developed countries occur in these countries (Bevan and Hood 2006; Heinrich and Marschke 2010; Jacob and Levitt 2003).

As Weberian states use performance management to reinvigorate states, public organizations in some sectors such as public education in Denmark have extensively implemented elements of performance management such as goal setting, performance tracking, attaching consequences to performance, and managerial autonomy (Hvidman and Andersen 2014; Holm 2018; Nielsen 2014b; Jacobsen and Andersen 2014). With few exceptions (Koning and Heinrich 2013), most research focuses on the impact of performance management on high-stake organizational performance, leaving the possible performance perversity largely unexamined. Without further empirical studies, it is hard to say whether the absence of studies on gaming can constitute evidence of the absence of gaming in the implementation of

performance management in these countries. Hence, more studies on the possible gaming behaviors in the implementation of performance management in these countries are needed to enrich our understanding of how political and environmental environment shapes or failed to shape the relationship between performance management and gaming.

Table 8. A Checklist Based on Contingency-Understanding of Performance Management to Evaluate the Likelihood of Performance Management Success

Dimension of Fit	Specific Questions
Horizontal Fit: Are performance management practices coherent?	(1) Are managers empowered to manage for results in exchange for accountability to results? (2) Is accurate performance feedback that meaningfully distinguishes between outstanding and mediocre employees/units provided to make possible learning from the past performance? (3) Do the public employees have the incentive and support to use performance information?
Vertical Fit: Are the organizational strategy and performance management practices aligned with the stage of performance management?	(1) Are sufficient resources and support allocated to assist the implementation of performance management?

	(2) At what stage of performance management, is the organization at (1) strategic planning and & establishment of the performance measurement system connected to the plan (2) implementing the changes in the plan and monitoring progress in achieving goals (3) performance review and preparation of the plan for next cycle. (3) Does the overall organizational strategy align with the stage of performance management? (4) If organizations fail to meet the goals for outputs or outcomes, does it revisit the logic model and understand what went wrong?
Organizational fit: Are the organizational resources including human capital sufficient to handle the tasks associated with fully implementing performance management?	(1) Are capable and ethical public employees in charge of implementing the reforms? (2) Do organizations have the resources to buffer the changes associated with the reforms?

	(3) Do organizations have a supportive environment for the implementation of performance management where underperformers are supported and overachievers are celebrated.
Environmental fit: Are the political and institutional environments compatible with the requirements of performance management?	This is still an active area of research.

APPENDICES

Appendix A

Table A-1. Summary statistics of variables at school level

Variable	N	Mean	Sd	Min	P50	Max
Setting standard	570	2.860	1.342	0	3	4
Teacher evaluation	570	1.026	0.602	0	1	3
Attach consequence to students' performance	570	4.123	3.707	0	5	11
Attach consequence to teachers' performance	570	0.601	0.714	0	0	3
Frequency of math test	570	1.058	1.296	0	1	6
Principal's autonomy	570	16.79	7.992	0	20	24
Test-Oriented Component	570	-0.0720	0.965	-2.183	0.0740	2.299
Personnel-Oriented Component	570	0.191	0.966	-1.336	0.447	2.516
School Socioeconomic score	570	-0.0310	0.385	-1.073	-0.0580	1.140
Principal's Competency	470	3.184	0.379	1.842	3.191	3.610
School relationship with stakeholders	470	1.541	0.577	1.034	1.248	3.504
State-centered school average math score in 2002	570	-1.309	4.795	-16.68	-1.262	16.41

Data Source: Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics.
Note. To comply with the requirements of using restricted dataset of ELS:2002, all the unweighted sample sizes were rounded to the nearest 10. Unless specified otherwise, the presentation of sample size follows the same rule.

Table A-2. Summary Statistics of Variables at Student Level

variable	N	mean	sd	min	p50	max
Score in 2004 Math tests	8830	49.86	10.09	19.82	49.84	76.52
Score in 2002 Math tests	8830	50.98	9.849	19.94	51.11	86.68
Gender	8800	0.487	0.500	0	0	1
Socioeconomic status	8800	0	0.606	-2.259	-0.00300	2.194
Black	8830	0.133	0.340	0	0	1
Hispanic	8830	0.141	0.348	0	0	1
Native American	8830	0.0100	0.0980	0	0	1
White	8830	0.561	0.496	0	1	1
Mixed	8830	0.0430	0.204	0	0	1
Asian	8830	0.112	0.315	0	0	1
Learning English as a second language	8300	0.0860	0.280	0	0	1
Special education	8340	0.0770	0.266	0	0	1
Hours spent on math homework	7170	5.650	3.079	2	6	18
Years of math-related coursework	8830	3.657	1.398	0	4	14
The teaching experience of the student's math teacher	7420	9.762	8.862	0	6	40

Data Source: Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics.
Note. To comply with the requirements of using restricted dataset of ELS:2002, all the unweighted sample sizes were rounded to the nearest 10. Unless specified otherwise, the presentation of sample size follows the same rule.

Appendix B
The Steps of Implementing Inverse Probability Weighting

In the first step, we used Twange package to implement the propensity score matching for schools in different groups. The variables used in the model include not only all the variables used in the main model, but also additional measures of organizational capacity such as the share of certified full-time and the share of part-time teachers, the average student perception of school safety, the location of schools (urban, suburban, and rural). Through the gradient boosting method, the propensity score implemented in Twang package takes into account the possibility of interaction terms and variables of higher powers (Ridgeway et al. 2017). The second step is to use the propensity score to construct the weight in the regression estimate.

As the study aims to investigate the difference in performances between schools in different levels of performance management, the study chose the average treatment effect (ATE) to generate the weight in the package. All the school-level control variables in the final model (Managerial emphasis on performance, the share of Hispanic, black students, and shares of male students, the lagged school average score) plus a series of variables (see table B-2) were used to create the propensity score of being in the specific group to which a school belongs.

For ATE, the weight for observations in a group is the inverse of the propensity score of being in that group (Ridgeway et al. 2017). Under the current weighting scheme, this method ensures that schools in each group that resemble those in the other groups get a larger weight. After weighing the observations according to the weight generated by the matching, the differences in groups became insignificant (see table B-3 and B-4). In addition, there was an overlap of sample distributions of estimated propensities generated by POC or TOC among groups and insignificant differences in covariate among groups (see figure B-1 and B-2). As no significant differences between the three groups of schools in school-level variables, we did not include the school-level variables in the multilevel modeling in models using IPW (Mccaffrey et al. 2013).

Table B-1.The list of variables

Dimension	Description	How it is constructed	Note
Dependent Variable	Math Score in 2004	norm-referenced measurements of	
Individual-level control variables	Math score in 2002	School-Centered measure of original data	
	Total years of math-related courses	School-Centered measure of students' self-reported year of math-related courses	
	Student's gender	Gender dummy	The responses are binary: 1. Male 0. Female
	Student's race	Dummy variables for Black, Hispanic Native American and mixed race students	Exclusion of Asian Americans and Pacific Islanders from minority status because they usually had the highest average math score (Barton 2009).
	Ever participating in English as a second language program	The response to the question of whether the student is ever in the program learning English as a second language	The responses were binary: 1.Yes 0.No
	Ever participating in a special education program	The response to the question of whether the student is ever in a special education program	The responses were binary: 1.Yes 0.No
	A student family's socioeconomic score	a School-centered variable of an equally weighted sum of parents'(guardians)education, occupations and family income.	
	The teacher's experience	Total years of teaching math at the school	
School-level Independent Variables of Interest	Test-Oriented Component	The first component score of principal component analysis on six types of measures of performance management.	.

		The Eigenvalue of this component was 2.647 and accounts for 44.1% of the total variance	
	Personnel-Oriented Component	The second component score of principal component analysis on six types of measures of performance management. The Eigenvalue of this component is 1.320 and accounts for 22.0% of the total variance.	
School past performance	The school 2002 performance		
The Social composition of schools	The school share of male students		
	The school share of African American and Hispanic students		
	The school average socioeconomic score	The average of students' socioeconomic scores within a school	

Data Source: Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics.

Table B-2. The Additional Variables in the Model for Propensity Score Matching

Dimension	How it is constructed	Note
The school safety	The school average of students' first-factor score on their perception of school safety issues in 2002. The first factor had an eigenvector of 3.936 and was the only factor with a larger than one eigenvalue.	The questions are how often the following event occurs to students Had something stolen at school Someone offered drugs at school Someone threatened to hurt 10th grader at school Got into a physical fight at school Someone hit 10th grader Someone forced money/things from 10th grader Someone damaged belongings Someone bullied or picked on 10th grader The response are ordinal: 1. Never 2. Once or twice 3. More than twice
Personnel's qualification	Percentage of certified full-time teachers in 2002	
	Percentage of certified part-time teachers in 2002	
School Location		The responses are categorical: 1; Urban 2: Suburban 3: Rural
State previous experience of implementing test-based accountability	The type of accountability implemented before 2002	The coding is based on Hanushek and Raymond (2005)'s categorization. 0:No accountability 1: report-card accountability 2: consequential accountability
	The years of implementing accountability till 2002	The coding is based on Dee and Jacob's (2011) work.
State implementation of subgroup student	The degree of requiring student subgroup performance on school report card	The codings for the following two item are based on Education Week's report (2003, pp. 86–88) .

accountability in the academic year 2002-2003		States gets four out of four if it requires schools to report disaggregated school performance by race, by poverty, by limited English Proficiency Student, and by special education students. For each category missing, one point is deducted.
	The degree of attaching consequences to school performance	States get seven out of seven, if its accountability includes (1) sanctions and sanctions contain the following options (2) school closure, (3) school reconstitution, (4) reconstitution as charter school, (5) permitting student transfer (6) transferring the management to a private entity (7) withholding funds
	The degree o applying incentive to school performance	The sanctions score plus one if a state provides rewards to high performing or improved schools.
State educational resources	State current educational expenditure per pupil	The value is log transformation of average state current educational expenditure between 2001-2002 and 2002 to 2003 academic year. As schools can not decide state average educational expenditure. We hold that this variable is exogenous.

Data Source: Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics.

Table B-3.The difference between schools in the three groups of POC after matching

Group1	Group2	Variable	Group1	Group2	Standardized Diff	P >|T|
Low	Middle	School_Math_02	50.62	50.93	0.06	0.55
Low	Middle	School_black_%	0.15	0.12	0.13	0.21
Low	Middle	School_hispanic_%	0.15	0.14	0.02	0.87
Low	Middle	School_male_%	0.47	0.49	0.12	0.27
Low	Middle	School_social_economic_score	-0.05	-0.03	0.05	0.64
Low	Middle	School_safety	1.49	1.49	0.02	0.84
Low	Middle	Certified_full_time teachers_%	96.67	97.55	0.08	0.3
Low	Middle	Certified_part_time teachers_%	65.14	72.43	0.16	0.2
Low	Middle	Location in urban area	0.29	0.26	0.06	0.85
Low	Middle	Location in suburban area	0.48	0.49	0.02	0.85
Low	Middle	Location in rural area	0.23	0.25	0.04	0.85
Low	Middle	Accountability: None before NCLB	0.13	0.14	0.02	0.96
Low	Middle	Report_Card_accountability before NCLB	0.11	0.11	0.02	0.96
Low	Middle	Consequential accountability before NCLB	0.76	0.74	0.03	0.96
Low	Middle	Year of implementing accountability	3.72	3.95	0.08	0.5
Low	Middle	Reporting of subgroup performance	3.33	3.22	0.07	0.55
Low	Middle	Attaching consequences to subgroup performance	2.11	2.03	0.04	0.73
Low	Middle	Log transformation of educational expenditure	9.1	9.1	0.01	0.92
Low	High	School_Math_02	50.62	50.66	0.01	0.94
Low	High	School_black_%	0.15	0.14	0.06	0.56
Low	High	School_hispanic_%	0.15	0.15	0	0.99
Low	High	School_male_%	0.47	0.49	0.13	0.19
Low	High	School_social_economic_score	-0.05	-0.01	0.09	0.35
Low	High	School_safety	1.49	1.49	0.01	0.96
Low	High	Certified_full_time teachers_%	96.67	96.59	0.01	0.94
Low	High	Certified_part_time teachers_%	65.14	70.32	0.11	0.35
Low	High	Location in urban area	0.29	0.24	0.11	0.61

Low	High	Location in suburban area	0.48	0.53	0.1	0.61
Low	High	Location in rural area	0.23	0.23	0.01	0.61
Low	High	Accountability: None before NCLB	0.13	0.11	0.08	0.3
Low	High	Report_Card_accountability before NCLB	0.11	0.16	0.15	0.3
Low	High	Consequential accountability before NCLB	0.76	0.73	0.06	0.3
Low	High	Year of implementing accountability	3.72	3.99	0.09	0.4
Low	High	Reporting of subgroup performance	3.33	3.17	0.1	0.39
Low	High	Attaching consequences to subgroup performance	2.11	2.01	0.05	0.67
Low	High	Log transformation of educational expenditure	9.1	9.09	0.08	0.48
Middle	High	School_Math_02	50.93	50.66	0.05	0.59
Middle	High	School_black_%	0.12	0.14	0.07	0.49
Middle	High	School_hispanic_%	0.14	0.15	0.02	0.89
Middle	High	School_male_%	0.49	0.49	0.02	0.86
Middle	High	School_social_economic_score	-0.03	-0.01	0.05	0.65
Middle	High	School_safety	1.49	1.49	0.03	0.79
Middle	High	Certified_full_time teachers_%	97.55	96.59	0.09	0.33
Middle	High	Certified_part_time teachers_%	72.43	70.32	0.05	0.68
Middle	High	Location in urban area	0.26	0.24	0.04	0.77
Middle	High	Location in suburban area	0.49	0.53	0.08	0.77
Middle	High	Location in rural area	0.25	0.23	0.05	0.77
Middle	High	Accountability: None before NCLB	0.14	0.11	0.11	0.27
Middle	High	Report_Card_accountability before NCLB	0.11	0.16	0.14	0.27
Middle	High	Consequential accountability before NCLB	0.74	0.73	0.03	0.27
Middle	High	Year of implementing accountability	3.95	3.99	0.01	0.89
Middle	High	Reporting of subgroup performance	3.22	3.17	0.03	0.77
Middle	High	Attaching consequences to subgroup performance	2.03	2.01	0.01	0.93
Middle	High	Log transformation of educational expenditure	9.1	9.09	0.07	0.54

Data Source: Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics.

Table B-4.The difference between schools in the three groups of TOC after matching

Group1	Group2	Variable	Group1	Group2	Standardized Diff	P >|T|
Low	Middle	School_Math_02	50.506	51.17	0.132	0.187
Low	Middle	School_black_%	0.149	0.117	0.14	0.135
Low	Middle	School_hispanic_%	0.137	0.14	0.012	0.91
Low	Middle	School_male_%	0.482	0.482	0.001	0.996
Low	Middle	School_social_economic_score	-0.009	-0.011	0.008	0.942
Low	Middle	School_safety	1.498	1.499	0.002	0.988
Low	Middle	Certified_full_time teachers_%	96.742	97.851	0.105	0.17
Low	Middle	Certified_part_time teachers_%	62.69	71.587	0.197	0.113
Low	Middle	Location in urban area	0.313	0.202	0.255	0.049
Low	Middle	Location in suburban area	0.502	0.551	0.099	0.049
Low	Middle	Location in rural area	0.186	0.247	0.143	0.049
Low	Middle	Accountability: None before NCLB	0.139	0.112	0.083	0.333
Low	Middle	Report_Card_accountability before NCLB	0.173	0.13	0.145	0.333
Low	Middle	Consequential accountability before NCLB	0.687	0.759	0.17	0.333
Low	Middle	Year of implementing accountability	4.183	3.86	0.111	0.307
Low	Middle	Reporting of subgroup performance	3.042	3.339	0.181	0.086
Low	Middle	Attaching consequences to subgroup performance	1.756	2.13	0.167	0.119
Low	Middle	Log transformation of educational expenditure	9.095	9.085	0.056	0.594
Low	High	School_Math_02	50.506	50.759	0.05	0.608
Low	High	School_black_%	0.149	0.155	0.023	0.82
Low	High	School_hispanic_%	0.137	0.16	0.102	0.332
Low	High	School_male_%	0.482	0.483	0.009	0.931
Low	High	School_social_economic_score	-0.009	-0.048	0.103	0.336
Low	High	School_safety	1.498	1.491	0.065	0.557
Low	High	Certified_full_time teachers_%	96.742	96.4	0.033	0.731
Low	High	Certified_part_time teachers_%	62.69	71.379	0.192	0.122
Low	High	Location in urban area	0.313	0.272	0.093	0.322
Low	High	Location in suburban area	0.502	0.477	0.05	0.322
Low	High	Location in rural area	0.186	0.251	0.152	0.322

Low	High	Accountability: None before NCLB	0.139	0.132	0.023	0.12
Low	High	Report_Card_accountability before NCLB	0.173	0.096	0.257	0.12
Low	High	Consequential accountability before NCLB	0.687	0.773	0.203	0.12
Low	High	Year of implementing accountability	4.183	3.454	0.25	0.024
Low	High	Reporting of subgroup performance	3.042	3.357	0.192	0.082
Low	High	Attaching consequences to subgroup performance	1.756	2.435	0.304	0.006
Low	High	Log transformation of educational expenditure	9.095	9.08	0.083	0.429
Middle	High	School_Math_02	51.17	50.759	0.082	0.421
Middle	High	School_black_%	0.117	0.155	0.163	0.081
Middle	High	School_hispanic_%	0.14	0.16	0.089	0.392
Middle	High	School_male_%	0.482	0.483	0.009	0.932
Middle	High	School_social_economic_score	-0.011	-0.048	0.095	0.367
Middle	High	School_safety	1.499	1.491	0.067	0.548
Middle	High	Certified_full_time teachers_%	97.851	96.4	0.138	0.093
Middle	High	Certified_part_time teachers_%	71.587	71.379	0.005	0.967
Middle	High	Location in urban area	0.202	0.272	0.162	0.259
Middle	High	Location in suburban area	0.551	0.477	0.149	0.259
Middle	High	Location in rural area	0.247	0.251	0.01	0.259
Middle	High	Accountability: None before NCLB	0.112	0.132	0.059	0.572
Middle	High	Report_Card_accountability before NCLB	0.13	0.096	0.112	0.572
Middle	High	Consequential accountability before NCLB	0.759	0.773	0.033	0.572
Middle	High	Year of implementing accountability	3.86	3.454	0.139	0.205
Middle	High	Reporting of subgroup performance	3.339	3.357	0.011	0.917
Middle	High	Attaching consequences to subgroup performance	2.13	2.435	0.137	0.213
Middle	High	Log transformation of educational expenditure	9.085	9.08	0.026	0.801

Data Source: Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics.

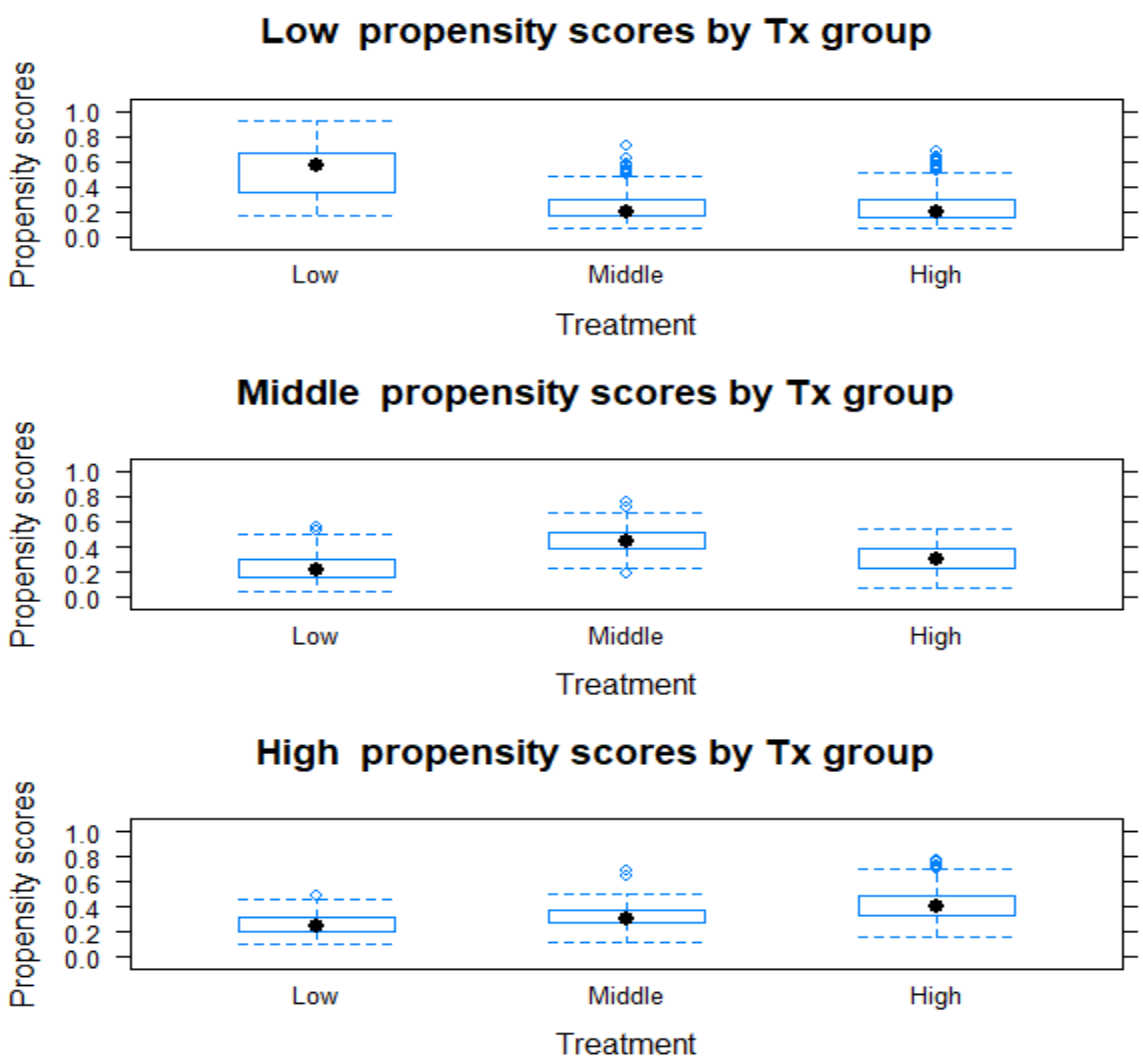

Figure B-1. The Distribution of Propensity Scores for groups of schools based on POC

Data Source: Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics.

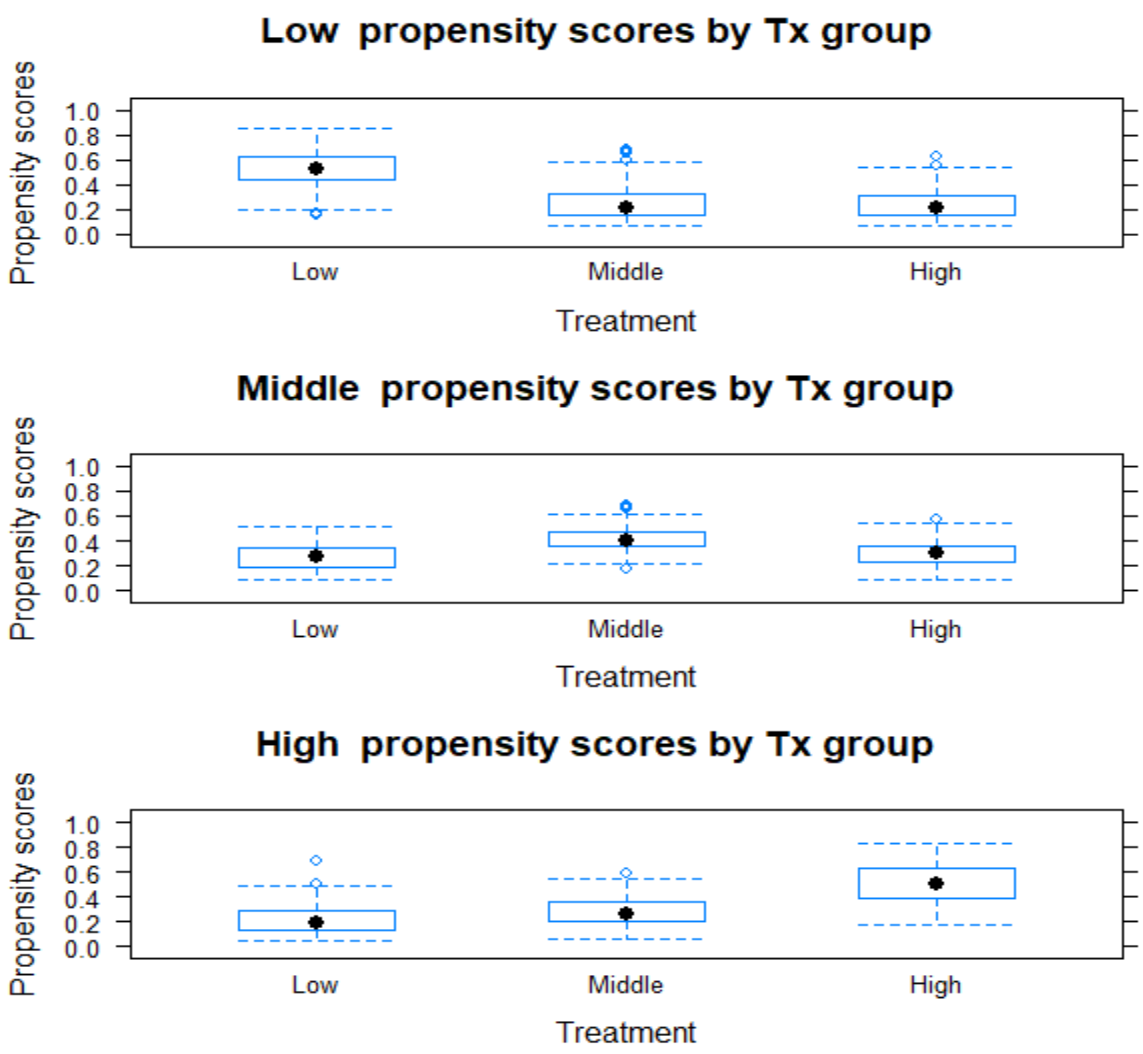

Figure B-2. The Distribution of Propensity Scores for groups of schools based on TOC

Data Source: Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics.

Appendix C

Table C-1. Abridged results after controlling for between-state variations

VARIABLES	Institutions	Instituions2	Capacity	All
Test-Oriented Component (TOC)	0.124	0.0802	0.0128	0.0186
	(0.140)	(0.143)	(0.138)	(0.139)
Test-Oriented Component^2 (TOC^2)	-0.112	-0.109	-0.0828	-0.0529
	(0.125)	(0.127)	(0.124)	(0.121)
Personnel-Oriented Component (POC)	0.337**	0.345**	0.386***	0.413***
	(0.134)	(0.136)	(0.134)	(0.132)
Personnel-Oriented Component (POC^2)	0.146*	0.143*	0.152**	0.149**
	(0.0757)	(0.0771)	(0.0750)	(0.0735)
TOC* State-Centered Math Score below the Mean^2	0.00177***	0.00176***	0.00180***	0.00180***
	(0.000606)	(0.000607)	(0.000606)	(0.000605)
POC* State-Centered Math Score below the Mean^2	-0.00350***	-0.00349***	-0.00349***	-0.00350***
	(0.000623)	(0.000624)	(0.000623)	(0.000622)
TOC* State-Centered Math Score above the Mean^2	0.000656	0.000646	0.000628	0.000623
	(0.000656)	(0.000657)	(0.000656)	(0.000655)
POC* State-Centered Math Score above the Mean^2	-0.000922	-0.000926	-0.000894	-0.000908
	(0.000728)	(0.000729)	(0.000728)	(0.000727)
TOC* African American	-0.0691	-0.0534	-0.0422	-0.0580
	(0.212)	(0.213)	(0.212)	(0.211)
POC* African American	0.0616	0.0517	0.0890	0.0887
	(0.202)	(0.203)	(0.202)	(0.201)
TOC* Hispanic	0.0904	0.0714	0.0309	0.0397
	(0.201)	(0.201)	(0.201)	(0.200)
POC* Hispanic	-0.305	-0.293	-0.278	-0.285
	(0.213)	(0.214)	(0.213)	(0.212)
TOC* Native	-0.588	-0.576	-0.656	-0.629
	(0.820)	(0.822)	(0.820)	(0.817)
POC* Native	-0.230	-0.213	-0.243	-0.311
	(0.779)	(0.780)	(0.778)	(0.775)
TOC* Limited English Proficiency	0.361*	0.361*	0.353	0.362*
	(0.219)	(0.219)	(0.219)	(0.219)
POC* Limited English Proficiency	-0.0143	-0.0110	-0.0149	-0.0139

	(0.243)	(0.243)	(0.243)	(0.242)
TOC* Special Education	0.00673	0.0166	0.0167	-0.00460
	(0.266)	(0.266)	(0.266)	(0.266)
POC* Special Education	0.279	0.273	0.265	0.290
	(0.285)	(0.285)	(0.285)	(0.285)
TOC* Socialeconomic status	-0.111	-0.110	-0.110	-0.112
	(0.0977)	(0.0978)	(0.0977)	(0.0977)
POC*Socialeconomic status	-0.0426	-0.0407	-0.0421	-0.0447
	(0.104)	(0.104)	(0.104)	(0.104)
Log of average educational expenditure			2.194***	1.055**
			(0.428)	(0.472)
Report Card Accountability before NCLB			0.533	0.654**
			(0.332)	(0.330)
Consequential Accountability before NCLB			0.757***	1.182***
			(0.274)	(0.281)
Years of Accountability before 2002 before NCLB			-0.0674**	-0.120***
			(0.0309)	(0.0332)
State Requirement of Disclosing Student Subgroup Performance after NCLB	0.188***	0.169***		0.229***
	(0.0514)	(0.0525)		(0.0589)
Inclusion of sanctions in state accountability system after NCLB	-0.200***			-0.205***
	(0.0408)			(0.0427)
Inclusion of both sanctions and reward in state accountability system after NCLB		-0.0969**		
		(0.0388)		
Observations	6,950	6,950	6,950	6,950
Icc	0.249	0.251	0.247	0.244

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, Version 1, U.S. Department of Education, National Center for Education Statistics.
Note. * 0.1 **0.05 ***0.01 (two-sided tests), and the standard error was clustered at school level. To comply with the requirements of using restricted dataset of ELS:2002, all the unweighted sample sizes were rounded to the nearest 10. Unless specified otherwise, the presentation of sample size follows the same rule.

Table C-2.The Abridged results on the correlation between performance management and organizational changes to improve students' math scores

VARIABLES	Block schedule for academic subject in 2004	Minutes of BS in 2004	Percentage of students in remedial math program in 2002	Percentage of teachers having eight or more hours trainings on teaching to students with limited English proficiency in 2002
User-oriented component (UOC)	0.0218	0.193	1.176*	0.0243**
	(0.165)	(3.343)	(0.710)	(0.0112)
UOC^2	0.255*	5.776*	-0.537	-0.00663
	(0.149)	(3.034)	(0.734)	(0.00850)
Provider-oriented component (POC)	0.311**	6.102**	0.505	-0.00901
	(0.154)	(2.909)	(0.874)	(0.00849)
POC ^2	0.216**	4.604**	0.581	0.00360
	(0.103)	(1.945)	(0.743)	(0.00517)
State requires the disclosure of student subgroup information	-0.0361	-0.622	-0.291**	-0.000447
	(0.0246)	(0.524)	(0.138)	(0.00230)
Inclusion of Sanctions in state accountability system	-0.0686	-1.437	0.298	0.00653
	(0.0736)	(1.458)	(0.322)	(0.00599)
Years of accountability before NCLB	-0.0134	-0.417	0.0135	0.00470
	(0.0555)	(1.098)	(0.242)	(0.00416)
Report-Card accountability before NCLB	0.0658	1.231	0.0254	-0.00540
	(0.0423)	(0.870)	(0.169)	(0.00417)
Consequential accountability before NCLB	1.202***	25.61***	-3.112	-0.0172
	(0.444)	(8.284)	(1.936)	(0.0293)
log transformation of educational expenditure per student	0.922**	18.81***	-3.736**	0.0431
	(0.379)	(6.623)	(1.629)	(0.0286)
logexp2	-0.758	-24.94**	8.145**	-0.181***
	(0.585)	(10.52)	(3.751)	(0.0652)
Constant	7.356	278.3***	-51.31	1.679***
	(5.361)	(96.25)	(31.88)	(0.525)
Observations	510	500	460	710
R-squared		0.096	0.076	0.228

Note. * 0.1 **0.05 ***0.01 (two-sided tests), and the standard error was clustered at school districts defined by National Center for Education Statistic.

Note. The control variables are the same as the school-level variables in the main models. Logit model was used as the first model and regression using ordinary least square estimation was used in the second to fourth models. Only the percentage of students receiving remedial math and teaching receiving 8 eight hours of training on teaching to students with limited English proficiency in 2002 were available.

Appendix D

Table D-1: The Operationalization of Key Variables Related to Civic Participation

Dimension	Variables	Operationalization	Notes
Dependent variables	Participated in volunteer work in 2004	The response to whether the student performed unpaid volunteer/community service work	Responses are binary: 1: Yes 0: No
	The sum of types of volunteering work in 2004	The sum of the following dummy variables: • Volunteered with youth organization • Volunteered with school/community service organization • Volunteered with political club/organization • Volunteered with church/church related group • Volunteered with community center/social-action group • Volunteered with hospital/nursing home group • Volunteered with education organization	Responses are binary: 1: Yes 0: No

		• Volunteered with conservation/environmental group	
	Frequency of Performing Community Service	The response to the question of how often performs community services	The responses are ordinal and in 2004 were coded as: 0.Rarely or never 1. Less than once a week 2. Once or twice a week 3. Everyday or almost everyday In 2006, the responses were changed to 0.No 1.Less than once a month 2.At least once a month, but not weekly 3.At least once a week
Cognitive Abilities	Knowledge of and skills in math	2002 and 2004 Standardized Math Score	
	Knowledge of and skills in reading	2002 standardized reading score	

Civic Skills	Relative socioeconomic Status(CSES)	A student family's socioeconomic score is a variable of an equally weighted sum of parents'(guardians)education, occupations and family income. To measure the relative social status, each student's socioeconomic status score was subtracted from the school average.	
	Categories of extracurricular activities	The sum of the ordinal responses to the following questions: • Participated in intramural sports • Participated in interscholastic sports • Participated in school band or chorus • Participated in school play or musical • Participated in student government • Participated in academic honor society • Participated in school yearbook or newsarticle • Participated in school service clubs • Participated in school academic clubs	The responses are ordinal: 0. Did not participate 1. Participate Participate as leaders/capital/officer

		• Participated in school hobby clubs Participated in school vocational clubs	
	Time spent on extracurricular activities per week	The answer to the question on the number of hours spent on extracurricular activities	The responses are ordinal: 0 None 1 Less than 1hour/week 2 1-4 hours 3 5-9 hours 4 10-14 hours 5 15-19 hours 6 20-24 hours 7 25 or more hours/week
Citizenship Norms	Individual endorsement of civic norms	the first-factor score of perceived importance of helping others in community, supporting environmental causes, being an informed and active citizen and being patriotic. The first factor had an eigenvalue of 1.887 and was the only factor that has an eigenvalue larger than 1.	The responses are ordinal: 1 Not important 2 Somewhat important 3 Very important The same number of factors as the number of variables were retained. The following analyses followed the same rule.
	School civic norms	The school average of individual endorsement of civic norms.	

	Parents' participation in school activities	The first factor score of parent (1) belong to parent-teacher organization (PTA) (2) participate in PTA meetings (3) participate in PTA activities (4) act as a volunteer (5) belong to other organizations with parents from school in 2002. The first factor had an eigenvalue of 2.827 and was the only factor that has an eigenvalue larger than 1.	The response are binary 1. Yes. 0. No

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1,U.S. Department of Education, National Center for Education Statistics.

Appendix E

The theories of civic participation can be classified into three types: (1) civic participation as a rational choice (2) civic participation as a moral obligation (3) civic participation as a social interaction. The main body of the article has introduced variables related to the first two schools of thought. The models also took into account variables related to social interactions. Previous studies have justified and used the following measures of social interactions (Campbell 2006, 2009; McFarland and Thomas 2006; O'Toole and Meier 2011).

Table E-1. The List of Control Variables Related to Social Interactions

Dimension	Variables	Operationalization	Notes
Social Interactions	Ties with friends	The first factor scores of the frequency of talking with friends, the frequency of hanging out with friends and the value of having strong friendship. The responses are ordinal The first factor had an eigenvalue of 0.887. According to Jolliffe (1972), the eigenvalue of 0.7 or larger is an acceptable cut-off point for selecting the number of factors.	The responses are ordinal: 1. Rarely 2. Less than once a week 3. Once or twice a week 4. Everyday or almost everyday

	Ties with parents	The first-factor score of the frequency of discussing with parents a series of topics as the indicator of the strength of ties with parents. The first factor had an eigenvalue of 4.887 and was the only factor with the eigenvalue larger than 1.	The responses are the same as the previous ones.
	Racial diversity	One minus the sum of the square of shares of students with different races at a school.	
	The disparity in individual's citizenship norm.	Standard deviation of individual commitment to citizenship norm within a school.	
	School relationship with stakeholders.	The first-factor score of administrators' evaluation of school relationship with school board, central office and teachers' association. The first factor had an eigenvector of 1.990 and was the only factor with a larger than one eigenvalue.	The responses were ordinal: 1 Very cooperative 2 Cooperative 3 Somewhat cooperative 4 Not cooperative

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics.

Additional Control Variables

Previous studies have demonstrated the importance of the following variables (Campbell 2006; Feldman and Matjasko 2005; Musick and Wilson 2008; O'Toole and Meier 2011), so the models examining the association of the intermediary variables with a student's volunteering behaviors also controlled for these variables.

Table E-2. The List of Additional Control Variables for Estimating the Relationships between Intermediary Variables and Volunteering

Variables	Operationalization	Notes
Socioeconomic status	The same operationalization as the one in Appendix B.	
Race dummies	Include dummies for African American, Hispanic, Native American, and Mixed Race	
Gender	Gender dummy	The responses were binary: 1. male 0. female
Educational aspiration for college	The response to the question of how far a student thinks he/she will get in education in 2002.	The responses were recoded as binary: 1.College or beyond 0.Below College
School student composition	The school share of African American, Hispanic, students and percentage of students eligible for free or reduced-price lunch	
Sponsored community service	Whether school-sponsored community service program was offered	1Yes 0.No

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics.

In the estimation on the influence of performance management on a student's reading and math score,the I also included a set of common control variables (Barton 2009; Fan and Wolters 2014; Jacob 2005; Lee and Bryk 1989).

Table E-3. The List of Additional Control Variables for Estimating the Relationships between Performance Management and Test Scores

Variables	Operationalization	Notes
Race dummies	The same operationalization as above.	
Gender	The same operationalization as above.	
Socioeconomic status	The same operationalization as above.	
Years of math-related course work	The sum of math-related coursework in 2004. ELS:2002 did not collect data on years of English-related coursework.	Similar questions on reading are missing probably because it can be challenging to exhaust courses related to reading.
Years of teaching math/English of the individual student's teacher	Year of teaching math/English in the current school when the survey of participating student's teachers was conducted in 2002.	
School student composition	The same operationalization as above.	
School baseline English/math performance	The school average students' reading/Math test score on ELS:2002 in 2002.	

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics.

In the estimation on the influence of performance management on a student's participation in extracurricular activities, the I also included a set of common control variables (Feldman and Matjasko 2005; Putnam 2007).

Table E-4. The List of Additional Control Variables for Estimating the Relationships between Performance Management and Participation in Extracurricular Activities

Variables	Operationalization	Notes
Race dummies	The same operationalization as above.	
Gender	The same operationalization as above.	
Socioeconomic status	The same operationalization as above.	
Students' previous participation in extracurricular activities	The sum of types/frequency of participating in extracurricular activities in 2002.	
Hours spent on homework	The average hour spent on math homework every week in 2004. The sum of average hours spent on math and English was not significant, so it was replaced with hours spent on math homework.	
Educational aspiration	The same operationalization as above.	
Ties with friends	The same operationalization as above.	
Ties with parents	The same operationalization as above.	

Racial diversity	The same operationalization as above.	
Disparity in individual's citizenship norm.	The same operationalization as above.	
School safety	The school average of first factor of students' response to the following questions in 2002. Had something stolen at school Someone offered drugs at school Someone threatened to hurt 10th grader at school Got into a physical fight at school Someone hit 10th grader Someone forced money/things from 10th grader Someone damaged belongings Someone bullied or picked on 10th grader The first factor had an eigenvalue of 3.936 and was the only factor with eigenvalue larger than 1.	The response are ordinal: 1. Never 2. Once or twice 3. More than twice

School socioeconomic status	The same operationalization as above.	
Ratio of Teacher to Students	The same operationalization as above.	The data come from the Common Core of data and can not be perfectly matched with all the public schools in ELS:2002.

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1,U.S. Department of Education, National Center for Education Statistics.

Table E-5. The List of Additional Control Variables for Estimating the Relationships between Performance Management and Students' Endorsement of Citizenship Norms

Variables	Operationalization	Notes
Race dummies	The same operationalization as above.	
Gender	The same operationalization as above.	
Socioeconomic status	The same operationalization as above.	
A student's perception of welcome class climate	School average first-factor scores of student responses to the following questions in 2002. Teachers are interested in students Teachers praise effort In class often feels put down by teachers In class often feels put down by students The first factor had an eigen value of 1.347 and was the only factor with an eigenvalue larger than 1.	Responses are ordinal: 0. Strongly disagree 1. Disagree 2. Agree 3. Strongly agree
Parents' participation in school activities (role modeling)	The first factor score of parent (1) belong to parent-teacher organization (PTA) (2) participate in PTA meetings (3) participate in PTA activities (4) act as a volunteer (5) belong to other organizations with parents from school in 2002. The first factor had an eigenvalue of 2.827 and was the only factor that has an eigenvalue larger than 1.	The response are binary 1. Yes.0. No
School average citizenship norm in base period	School average perceived importance of helping others in the community in 2002.	Only this question used to construct factor score in citizenship norm in 2004 was asked in 2002 The responses are ordinal 1;No very important 2: Somewhat important 3: Very important

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics.

Appendix F

Table F-1. Summary Statistics of School-Level Variables

Variable	N	Mean	SD	Min	Median	Max
Setting Standard	570	2.860	1.342	0	3	4
Teacher Evaluation	570	1.026	0.602	0	1	3
Attach consequence to students' performance	570	4.123	3.707	0	5	11
Attach consequence to teachers' performance	570	0.601	0.714	0	0	3
Frequency of math test	570	1.058	1.296	0	1	6
Frequency of English test	570	1.003	1.229	0	1	6
Principal's autonomy	570	16.79	7.992	0	20	24
User-oriented component	570	0.169	0.967	-1.089	0.319	3.306
Provider-oriented component	570	-0.0400	0.982	-2.251	0.166	2.060
School-sponsored community service programs	510	0.777	0.416	0	1	1
Percentage of students eligible for free or reduced-priced lunch	380	0.295	0.203	0.00100	0.252	0.955
School percentage of Black students	390	0.168	0.220	0.00300	0.0670	0.998
School Percentage of Hispanic students	390	0.122	0.186	0.00300	0.0330	0.899
School Relationship with stakeholders	460	1.541	0.577	1.034	1.248	3.504
Racial Heterogeneity	570	0.612	0.217	0.219	0.590	1
School civic norms	550	0.0130	0.172	-0.502	0.0120	0.609
Standard deviation of School Civic Norms	570	0.0540	0.0110	0	0.0540	0.0890
School average perceived importance of helping others in the community in the base period (2002)	400	2.308	0.171	1.889	2.308	3

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics. (2) Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2000-2001, 2000-2001, Version 1, U.S. Department of Education, National Center for Education Statistics.

Note. To comply with the requirements of using restricted dataset of ELS:2002, all the unweighted sample sizes were rounded to the nearest 10.

Table F-2. Summary Statistics of Student-Level Variables

Variable	N	Mean	SD	Min	Median	Max
Participation in volunteering in 2002	8270	0.305	0.461	0	0	1
Participation in volunteering in 2004	8720	0.596	0.491	0	1	1
Participation in volunteering in 2006	7880	0.453	0.498	0	0	1
Frequency of volunteering in 2002	8270	0.434	0.732	0	0	3
Frequency of volunteering in 2004	8690	1.639	0.845	1	1	4
Frequency of volunteering in 2006	7870	0.834	1.060	0	0	3
Types of volunteering in 2004	8830	1.078	1.418	0	0	8
Types of volunteering in 2006	8830	0.808	1.292	0	0	8
Hours spent on extracurricular activities every week in 2002	8390	2.444	1.433	1	2	6
Hours spent on extracurricular activities every week in 2004	8790	2.130	1.921	0	2	7
Types of extracurricular activities in 2002	8830	2.110	2.136	0	2	26
Types of extracurricular activities in 2004	8830	1.889	2.089	0	1	18
Gender	8800	0.487	0.500	0	0	1
Black	8830	0.133	0.340	0	0	1
Hispanic	8830	0.141	0.348	0	0	1
White	8830	0.561	0.496	0	1	1
Asian	8830	0.112	0.315	0	0	1
Native	8830	0.0100	0.0980	0	0	1
School-centered individual socioeconomic score	8800	0	0.606	-2.259	-0.003	2.194
Score in 2004 Math test	8830	49.86	10.09	19.82	49.84	76.52
Score in 2002 Math test	8830	50.98	9.849	19.94	51.11	86.68
Years of math-related coursework	8830	3.657	1.398	0	4	14
The teaching experience of the student's math teacher	7420	9.762	8.862	0	6	40
Score in 2002 reading test	8830	50.48	9.796	22.57	50.79	78.76
ACT reading	3380	20.96	5.909	1	20	36
The teaching experience of the student's English teacher	7160	9.647	9.277	0	5	40
Student's perception of welcome class climate	8180	2.764	0.563	-0.257	2.810	3.934
Educational aspiration for colleges	8830	0.706	0.455	0	1	1
Commitment to civic norms	8590	2.523	0.540	1.128	2.457	3.384
Parents' participation in school activities	7030	0.319	0.370	0	0.157	1.157
Strength of ties with friends	8630	3.009	0.747	0.980	3.230	3.918

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics.
(2) Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2000-2001, 2000-2001, Version 1, U.S. Department of Education, National Center for Education Statistics.
Note. To comply with the requirements of using restricted dataset of ELS:2002, all the unweighted sample sizes were rounded to the nearest 10.

Appendix G

Reasons for and Methods of Controlling Response Styles

First, taking into account ARS and DARS can reduce the common source bias caused by students' personality and time pressure. An extrovert who tend to seek stimulation and accept statements impulsively are more likely to agree with items while an introvert who is inclined to stay away from external simulation is more likely to disagree with items (Couch and Keniston 1960). For these reasons, ARS and DARS can affect both self-reported volunteering behaviors, the dependent variable, and, self-reported participation in extracurricular activities, the intermediating variables. Failing to control for the two styles can distort the inference about the indirect effects of performance management on student volunteering behaviors through participation in extracurricular activities. I followed Baumgartner and Steenkamp's method (2001) and operationalized the ARS(DARS) by measuring the extent to which students agreed (disagreed) with similar items that are both positively and negatively worded and many heterogeneous items. The correlations among the heterogeneous items were less than 0.3 and indicated weak correlation among the items used to measure the response styles.

Second, controlling for NCR and ERS can reduce random measurement errors. Students were likely to respond stylistically to questions because they may not have sufficient motivation to carefully answer the survey questions. Before filling out the questionnaire, students took a 60-minutes standardized test of reading and math in 2002 and a 26-minutes standardized math test in 2004 (Ingels et al. 2005). The questionnaire contained 98 questions in 2002 and 64 questions in 2004 (RTI 2004; 2002). Many questions had multiple sub-questions. 2002 questionnaires ran as long as 38 pages while 2004 survey had 27 pages. Students needed to finish the survey during a 45 minutes group administration. Students could become fatigued by the challenging questions of math and reading and a long stream of questions of which the significanc was not clear to students (Podsakoff, MacKenzie, and Podsakoff 2012). It is possible that students may uncritically endorse or negate a series of correlated items so that they reduce fatigue or respond to the pressure for finishing the survey in time. NCR aims to capture this tendency.

Following Baumgartner and Steenkamp's method (2001), I generated NCR by summing up the absolute differences between pairs of closely related questions. Alternatively, students may just give extreme responses to all the questions because they wanted to finish the survey quickly or had less tolerance with ambiguity (Naemi, Beal, and Payne 1992). Following Baumgartner and Steenkamp's method (2001), I generated ERS by counting the time a respondent endorsed the extreme response to a series of heterogeneous items. Controlling for both ERS and NCR can lower the noise caused by students' random or extreme responses to questions. Less noise can also reduce the possibility of attenuated relationships between participation in extracurricular activities and volunteering behaviors (Podsakoff, MacKenzie, and Podsakoff 2012).

Coding Response Styles Acquiatance response style (ARS)-the tendency to agree with items regardless of content

Disacquiatance response style (DARS)-the tendency to disagree with items regardless of content.

Four item responses-strongly agree, agree, disagree, strongly disagree

ARS increases by 2 if the response is strongly agree; ARS increases by 1 if the response is agree.

DARS increases by 2 if the response is strongly disagree; DARS increases by 1 if the response is disagree.

Similar items that are both positively and negatively worded

Students friendly with other racial groups

Racial/ethnic groups often fight

The teaching is good

In class often feels put down by teachers

Heterogeneous Items- none of the correlations among the following variables are above 0.3, indicating a lack of commonality and meeting the requirement of heterogeneity. Pearson rather than polychoric correlation was used because of the failure of performing polychoric correlation in this case.

Has nothing better to do than school in 2002

There is real school spirit in 2002

Everyone knows the school rule in 2002

Does not feel safe at this school in 2002

Punishment same no matter who you are in 2002

Education is important to get a job later in 2002

Teachers expect success in school in 2002

Plays on a team or belongs to a club in 2002

Two item responses

ARS increases by one if the respondent agrees with the statement. DARS increases by one if the respondent disagrees with the statement.

Won an academic honor in 2002

Participated in voc/tech skills competitionin 2002

Uses computers in math class in 2002

Parents know 1st friend's parents in 2002

Uses computer in 10th grade spring social studies in 2002

Ever in career academy in 2002

Play an individual intramural sport in 2002

Plans to go on to school right after high school in 2004

Non-Contingent Response-the tendency to respond to items carelessly and randomly.

All of these pairs had polychoric correlation higher than 0.5 except the fourth pair (0.48). For the fourth pair, the T test showed that responses between the two variables were not significant even at 0.1 significance level.

In class often feels put down by teachers in 2002

In class often feels put down by students in 2002

School rules are fair in 2002

Punishment same no matter who you are in 2002

Won an academic honor in 2002

Recognized for good grades in 2002

How often listens to math teacher lecture in 2002

How often copies math teacher's notes from board in 2002

Gets totally absorbed in mathematics in 2002

Thinks math is fun in 2002

Gets totally absorbed in reading in 2002

Thinks reading is fun in 2002

Can do excellent job on math tests in 2004

Can do excellent job on math assignments in 2004

Can do excellent job on English tests in 2004

Can do excellent job on English assignments in 2004

Has gone to counselor for college entrance information

Has gone to teacher for college entrance information

Mother's desire for respondent after high school

Father's desire for respondent after high school

How far in school mother wants respondent to go

How far in school father wants respondent to go

Extreme Response Style (ERS)-the tendency of endorsing most extreme responses regardless of content.

If the response is equal to the minimal or maximum value of the variable, 1 is added to ERS. None of the correlations was above 0.2, indicating little commonality among the variables.

How often does problem-solving in math class in 2002

How often math teacher uses computer to instruct one-on-one in 2002

How often goes to class without homework done in 2002

How often volunteers or performs community service in 2002

How often drives or rides around in 2002

How often uses computer at another place in 2002

How often 10th grader speaks native language with friends in 2002

How often parents help with homework in 2002

How often use calculator in math class in 2004

How often uses computer at friend's house in 2004

How often takes music, art, language class in 2004

How often takes sports lessons in 2004

Use of public library for assignments in 2004

Appendix H

Table H-1. The Complete Results with ACT Reading Score as the Intermediary Variable

VARIABLES	ACT Reading	Math04	Extratype04	Extrahour04	Civic04	Volr04	Voltype04	Volfreq04
User-oriented component (UOC)	-0.0982	0.323**	-0.0322	0.00177	-0.00718			
	(0.148)	(0.141)	(0.0408)	(0.0321)	(0.0126)			
UOC^2	0.190*	-0.0966	0.00293	-0.0184	0.0101			
	(0.104)	(0.0934)	(0.0287)	(0.0206)	(0.00891)			
Provider-oriented component (POC)	-0.0620	0.0431	-0.0380	-0.0863***	0.0178*			
	(0.163)	(0.126)	(0.0342)	(0.0295)	(0.0107)			
POC^2	-0.0467	0.129	-0.0377*	-0.0292	0.00334			
	(0.101)	(0.0801)	(0.0217)	(0.0179)	(0.00710)			
Math score from ELS:2002						0.0383***	0.0134***	0.0279***
						(0.00817)	(0.00308)	(0.00642)
ACT reading score						0.0238*	0.00316	-0.0117
						(0.0124)	(0.00412)	(0.00885)
School-centered individual socioeconomic score	1.791***	3.406***	0.517***	0.407***	0.0314***	0.257***	0.0540*	0.0853
	(0.185)	(0.179)	(0.0389)	(0.0348)	(0.0122)	(0.0856)	(0.0303)	(0.0710)
Individual Sum of extracurricular activity types						0.334***	0.111***	0.215***
						(0.0380)	(0.00795)	(0.0218)
Individual Frequency of engaging in extracurricular activities						0.126***	0.0637***	0.0785***
						(0.0261)	(0.0101)	(0.0236)
Individual endorsement of civic norms						0.712***	0.308***	0.957***
						(0.103)	(0.0424)	(0.0835)
School civic norms						0.593*	0.0913	0.339
						(0.360)	(0.177)	(0.283)
School standard deviation of civic norms						5.649	-0.207	-0.138
						(5.804)	(2.253)	(5.038)
Strength of ties with friends			0.000360	0.175***		0.0345	0.0103	0.145**
			(0.0293)	(0.0272)		(0.0730)	(0.0275)	(0.0596)
School's relationships with stakeholders						0.00113	0.0140	0.0876
						(0.0914)	(0.0355)	(0.0814)
Whether the school offers service learning programs						0.395***	0.147**	0.259**
						(0.130)	(0.0594)	(0.113)
Student's intention of going to college			0.889***	1.042***		0.405***	0.231***	0.243

			(0.0451)	(0.0454)		(0.144)	(0.0752)	(0.151)
Asian	-0.225	1.287***	0.269***	-0.138**	0.0204	0.354*	0.114**	0.503***
	(0.368)	(0.350)	(0.0853)	(0.0594)	(0.0242)	(0.209)	(0.0563)	(0.147)
African American	-3.471***	-4.957***	-0.218**	0.206***	0.0555**	0.0955	0.215***	0.194
	(0.377)	(0.364)	(0.0869)	(0.0764)	(0.0258)	(0.202)	(0.0805)	(0.172)
Hispanic	-1.825***	-2.849***	-0.189**	-0.0120	0.0469*	-0.0776	-0.0506	0.0592
	(0.414)	(0.382)	(0.0793)	(0.0727)	(0.0259)	(0.223)	(0.0921)	(0.189)
Native American	-3.080***	-3.391**	-0.485**	0.123	0.0982	-0.160	0.285	-0.288
	(1.114)	(1.414)	(0.219)	(0.173)	(0.0801)	(0.526)	(0.182)	(0.530)
Mixed-race	-0.609	-0.998**	-0.0585	0.0951	-0.0782**	0.166	0.166*	0.0795
	(0.561)	(0.506)	(0.0999)	(0.104)	(0.0385)	(0.286)	(0.0950)	(0.189)
Male	-0.503**	1.087***	-0.711***	0.233***	-0.100***	-0.642***	-0.248***	-0.397***
	(0.215)	(0.197)	(0.0447)	(0.0409)	(0.0136)	(0.109)	(0.0390)	(0.0935)
Share of students eligible for reduced-price or free meals	-2.086***	-2.106***	0.187	0.736***	0.0150	-0.598*	-0.119	-0.525
	(0.668)	(0.677)	(0.165)	(0.152)	(0.0533)	(0.352)	(0.177)	(0.358)
Share of African American students	1.232*	4.748***	0.0811	-0.530***	0.0797	-0.450	-0.0679	-0.237
	(0.668)	(0.633)	(0.162)	(0.144)	(0.0491)	(0.377)	(0.163)	(0.312)
Share of Hispanic students	0.942	3.519***	-0.310*	-0.247	0.0799	0.249	0.0840	0.735*
	(0.863)	(0.695)	(0.170)	(0.161)	(0.0569)	(0.471)	(0.205)	(0.417)
School average reading score in 2002	0.372***							
	(0.0264)							
Students' English teacher experience	0.0454***							
	(0.0112)							
School average math score at base period		0.802***						
		(0.0271)						
The number of math-related courses		1.589***						
		(0.0881)						
Student's math teacher experience		0.0960***						
		(0.0137)						
School average sum of extracurricular activities types at base period			0.861***					
			(0.0568)					
School average frequency of engaging in extracurricular activities at base period				0.613***				
				(0.0466)				
Parents' participation in school activities					0.0679***			

					(0.0191)			
Students' perception of a welcome class environment					0.0838***			
					(0.0136)			
The school average perceived importance of helping others at base period					0.350***			
					(0.0479)			
Constant	2.357	2.124	0.784***	-0.803***	1.465***	-5.001***	-1.988***	
	(1.495)	(1.518)	(0.128)	(0.159)	(0.115)	(0.591)	(0.270)	
Number of observations used in the specific equation	2740	6950	8080	8060	5980	2450	2460	2440
Number of observations ever used in the whole model	8,260							

Robust standard errors in parentheses (two-sided tests)

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics.
(2) Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2000-2001, 2000-2001, Version 1, U.S. Department of Education, National Center for Education Statistics. The next nine tables use the same dataset.
Note. * 0.1 **0.05 ***0.01 (two-sided tests), and the standard error was clustered at school level. To comply with the requirements of using restricted dataset of ELS:2002, all the unweighted sample sizes were rounded to the nearest 10.

Table H-2. The Complete Results with Reading Score from ELS:2002 as the Intermediary Variable

VARIABLES	Read02	Math04	extratype04k	extrahour04	civic	volunteer04	voltype04	volfreq04
User-oriented component (UOC)	0.171**	0.323**	-0.0322	0.00177	-0.00718			
	(0.0861)	(0.141)	(0.0408)	(0.0321)	(0.0126)			
UOC^2	-0.0520	-0.0966	0.00293	-0.0184	0.0101			
	(0.0556)	(0.0934)	(0.0287)	(0.0206)	(0.00891)			
Provider-oriented component (POC)	-0.000142	0.0431	-0.0380	-0.0863***	0.0178*			
	(0.0759)	(0.126)	(0.0342)	(0.0295)	(0.0107)			
POC^2	0.112**	0.129	-0.0377*	-0.0292	0.00334			
	(0.0478)	(0.0801)	(0.0217)	(0.0179)	(0.00710)			
Math score from ELS:2002						0.0333***	0.0132***	0.0177***
						(0.00456)	(0.00229)	(0.00408)
Reading score from ELS: 2002						0.0207***	0.00454**	-0.00101
						(0.00485)	(0.00220)	(0.00402)
School-centered individual socioeconomic score	3.933***	3.406***	0.517***	0.407***	0.0314***	0.242***	0.0795***	0.0936*
	(0.181)	(0.179)	(0.0389)	(0.0348)	(0.0122)	(0.0553)	(0.0222)	(0.0478)
Individual Sum of extracurricular activity types						0.324***	0.112***	0.221***
						(0.0225)	(0.00722)	(0.0141)
Individual Frequency of engaging in extracurricular activities						0.125***	0.0713***	0.0965***
						(0.0174)	(0.00807)	(0.0156)
Individual endorsement of civic norms						0.433***	0.282***	0.828***
						(0.0587)	(0.0289)	(0.0542)
School civic norms						0.0129	0.110	0.0853
						(0.232)	(0.124)	(0.194)
School standard deviation of civic norms						-2.016	-1.012	-1.956
						(3.914)	(1.888)	(3.315)
Strength of ties with friends			0.000360	0.175***		0.0253	0.0236	0.118***
			(0.0293)	(0.0272)		(0.0418)	(0.0212)	(0.0361)
School's relationships with stakeholders						0.0588	0.0354	0.127**
						(0.0656)	(0.0280)	(0.0560)
Whether the school offers service learning programs						0.253***	0.141***	0.201***
						(0.0840)	(0.0435)	(0.0744)
Student's intention of going to college			0.889***	1.042***		0.426***	0.367***	0.370***
			(0.0451)	(0.0454)		(0.0776)	(0.0481)	(0.0695)
Asian	-0.938***	1.287***	0.269***	-0.138**	0.0204	0.346***	0.0785*	0.343***

	(0.307)	(0.350)	(0.0853)	(0.0594)	(0.0242)	(0.117)	(0.0441)	(0.0884)
African American	-4.305***	-4.957***	-0.218**	0.206***	0.0555**	0.179	0.126*	0.197*
	(0.376)	(0.364)	(0.0869)	(0.0764)	(0.0258)	(0.119)	(0.0671)	(0.112)
Hispanic	-3.020***	-2.849***	-0.189**	-0.0120	0.0469*	0.0187	-0.0599	0.0168
	(0.382)	(0.382)	(0.0793)	(0.0727)	(0.0259)	(0.114)	(0.0596)	(0.109)
Native American	-4.224***	-3.391**	-0.485**	0.123	0.0982	-0.423*	-0.112	0.0250
	(1.006)	(1.414)	(0.219)	(0.173)	(0.0801)	(0.228)	(0.180)	(0.268)
Mixed-race	-1.165**	-0.998**	-0.0585	0.0951	-0.0782**	0.212	0.192***	0.0516
	(0.517)	(0.506)	(0.0999)	(0.104)	(0.0385)	(0.166)	(0.0684)	(0.132)
Male	-1.484***	1.087***	-0.711***	0.233***	-0.100***	-0.684***	-0.318***	-0.326***
	(0.202)	(0.197)	(0.0447)	(0.0409)	(0.0136)	(0.0664)	(0.0298)	(0.0579)
Share of students eligible for reduced-price or free meals	-0.533	-2.106***	0.187	0.736***	0.0150	-0.147	-0.0451	-0.200
	(0.365)	(0.677)	(0.165)	(0.152)	(0.0533)	(0.223)	(0.127)	(0.228)
Share of African American students	3.787***	4.748***	0.0811	-0.530***	0.0797	-0.407*	0.0775	0.00648
	(0.463)	(0.633)	(0.162)	(0.144)	(0.0491)	(0.243)	(0.126)	(0.238)
Share of Hispanic students	2.925***	3.519***	-0.310*	-0.247	0.0799	0.197	0.0798	0.381
	(0.536)	(0.695)	(0.170)	(0.161)	(0.0569)	(0.283)	(0.135)	(0.235)
School average reading score in 2002	0.965***							
	(0.0139)							
Students' English teacher experience	0.0594***							
	(0.0116)							
School average math score at base period		0.802***						
		(0.0271)						
The number of math-related courses		1.589***						
		(0.0881)						
Student's math teacher experience		0.0960***						
		(0.0137)						
School average sum of extracurricular activities types at base period			0.861***					
			(0.0568)					
School average frequency of engaging in extracurricular activities at base period				0.613***				
				(0.0466)				
Parents' participation in school activities					0.0679***			
					(0.0191)			
Students' perception of a welcome class environment					0.0838***			

					(0.0136)			
The school average perceived importance of helping others at base period					0.350***			
					(0.0479)			
Constant	2.281***	2.124	0.784***	-0.803***	1.465***	-4.346***	-2.327***	
	(0.792)	(1.518)	(0.128)	(0.159)	(0.115)	(0.401)	(0.201)	
Number of observations used in the specific equation	6740	6950	8080	8060	5960	5960	6020	5960
Number of observations ever used in the whole model	8270							

Robust standard errors in parentheses (two-sided tests)

*** p<0.01, ** p<0.05, * p<0.1

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics.
(2) Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2000-2001, 2000-2001, Version 1, U.S. Department of Education, National Center for Education Statistics. The next nine tables use the same dataset.

Table H-3. The Abridged Results for Volunteering Behaviors in Next Period after Controlling for Volunteering Behaviors at Base Level (ACT reading scores as the Intermediary Variable)

Variable	ACT reading	Math04	Extratype04	Extrahour04	Civic	Volunteer06	Voltype06	Volfreq06
User-oriented component (UOC)	-0.0982	0.323**	-0.0322	0.00177	-0.00718			
	(0.148)	(0.141)	(0.0408)	(0.0321)	(0.0126)			
UOC^2	0.190*	-0.0966	0.00293	-0.0184	0.0101			
	(0.104)	(0.0934)	(0.0287)	(0.0206)	(0.00891)			
Provider-oriented component (POC)	-0.0620	0.0431	-0.0380	-0.0863***	0.0178*			
	(0.163)	(0.126)	(0.0342)	(0.0295)	(0.0107)			
POC^2	-0.0467	0.129	-0.0377*	-0.0292	0.00334			
	(0.101)	(0.0801)	(0.0217)	(0.0179)	(0.00710)			
Math score from ELS:2002						0.0273***	0.0245***	0.0163***
						(0.00743)	(0.00641)	(0.00421)
ACT reading score						0.0214**	0.0182**	0.0110*
						(0.00986)	(0.00826)	(0.00571)
School-centered individual socioeconomic score	1.791***	3.406***	0.517***	0.407***	0.0314***	0.195**	0.152**	0.0955**
	(0.185)	(0.179)	(0.0389)	(0.0348)	(0.0122)	(0.0790)	(0.0707)	(0.0456)
Individual Sum of extracurricular activity types						0.147***	0.0934***	0.0915***
						(0.0273)	(0.0186)	(0.0115)
Individual Frequency of engaging in extracurricular activities						0.125***	0.110***	0.0639***
						(0.0262)	(0.0227)	(0.0149)
Individual commitment to civic norms						0.495***	0.476***	0.351***
						(0.0934)	(0.0821)	(0.0501)
volunteering in base period						0.543***		0.426***
						(0.0971)		(0.0503)
volunteering frequency in 2002							0.467***	
							(0.0607)	
Constant	2.357	2.124	0.784***	-0.803***	1.465***	-4.022***		-2.651***
	(1.495)	(1.518)	(0.128)	(0.159)	(0.115)	(0.614)		(0.350)
Number of observations used in the specific equation	2740	6950	8080	8060	5980	2450	2460	2440
Number of observations ever used in the whole model	8,260							

Robust standard errors in parentheses (two-sided tests)

*** p<0.01, ** p<0.05, * p<0.1

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics.

(2) Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2000-2001, 2000-2001, Version 1, U.S. Department of Education, National Center for Education Statistics. The next nine tables use the same dataset.

Table H-4. The Abridged Results for Volunteering Behaviors in the Next Period after Controlling for Volunteering Behaviors at Base Level (Reading Scores from ELS:2002 as the Intermediary Variable)

VARIABLES	Read02	Math04	extratype04k	extrahour04	civic	volunteer06	volfreq06	voltype06
User-oriented component (UOC)	0.171**	0.323**	-0.0322	0.00177	-0.00718			
	(0.0861)	(0.141)	(0.0408)	(0.0321)	(0.0126)			
UOC^2	-0.0520	-0.0966	0.00293	-0.0184	0.0101			
	(0.0556)	(0.0934)	(0.0287)	(0.0206)	(0.00891)			
Provider-oriented component (POC)	-0.000142	0.0431	-0.0380	-0.0863***	0.0178*			
	(0.0759)	(0.126)	(0.0342)	(0.0295)	(0.0107)			
POC^2	0.112**	0.129	-0.0377*	-0.0292	0.00334			
	(0.0478)	(0.0801)	(0.0217)	(0.0179)	(0.00710)			
Math score from ELS:2002						0.0285***	0.0250***	0.0188***
						(0.00482)	(0.00452)	(0.00320)
Reading score from ELS: 2002						0.00731	0.00345	0.00226
						(0.00452)	(0.00429)	(0.00311)
School-centered individual socioeconomic score	3.933***	3.406***	0.517***	0.407***	0.0314***	0.211***	0.162***	0.108***
	(0.181)	(0.179)	(0.0389)	(0.0348)	(0.0122)	(0.0555)	(0.0512)	(0.0376)
Individual Sum of extracurricular activity types						0.156***	0.118***	0.108***
						(0.0196)	(0.0143)	(0.00964)
Individual Frequency of engaging in extracurricular activities						0.118***	0.108***	0.0716***
						(0.0182)	(0.0160)	(0.0120)
Individual commitment to civic norms						0.408***	0.393***	0.355***
						(0.0590)	(0.0559)	(0.0412)
volunteering in base period						0.560***		0.433***
						(0.0689)		(0.0398)
volunteering frequency in 2002							0.404***	
							(0.0412)	
Constant	2.281***	2.124	0.784***	-0.803***	1.465***	-3.753***		-2.784***
	(0.792)	(1.518)	(0.128)	(0.159)	(0.115)	(0.401)		(0.265)
Number of observations used in the specific equation	6740	6950	8080	8060	5980	5960	6020	5960
Number of observations ever used in the whole model	8270							

Robust standard errors in parentheses (two-sided tests)

*** p<0.01, ** p<0.05, * p<0.1

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics.

(2) Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2000-2001, 2000-2001, Version 1, U.S. Department of Education, National Center for Education Statistics. The next nine tables use the same dataset.

Table H-5 The Abridged Results for Volunteering Behaviors after Controlling for State Fixed Effect (ACT Reading Score as the Intermediary Variable)

VARIABLES	ACT reading	Math04	Extratype04	Extrahour04	Civic	Volunteer04	Voltype04	Volfreq04
User-oriented component (UOC)	-0.0982	0.323**	-0.0322	0.00177	-0.00718			
	(0.148)	(0.141)	(0.0408)	(0.0321)	(0.0126)			
UOC^2	0.190*	-0.0966	0.00293	-0.0184	0.0101			
	(0.104)	(0.0934)	(0.0287)	(0.0206)	(0.00891)			
Provider-oriented component (POC)	-0.0620	0.0431	-0.0380	-0.0863***	0.0178*			
	(0.163)	(0.126)	(0.0342)	(0.0295)	(0.0107)			
POC^2	-0.0467	0.129	-0.0377*	-0.0292	0.00334			
	(0.101)	(0.0801)	(0.0217)	(0.0179)	(0.00710)			
Math score from ELS:2002						0.0344***	0.0118***	0.0268***
						(0.00869)	(0.00312)	(0.00680)
ACT reading score						0.0247*	0.00368	-0.00970
						(0.0128)	(0.00411)	(0.00913)
School-centered individual socioeconomic score	1.791***	3.406***	0.517***	0.407***	0.0314***	0.243***	0.0489	0.0588
	(0.185)	(0.179)	(0.0389)	(0.0348)	(0.0122)	(0.0878)	(0.0304)	(0.0731)
Individual Sum of extracurricular activity types						0.340***	0.111***	0.222***
						(0.0397)	(0.00817)	(0.0223)
Individual Frequency of engaging in extracurricular activities						0.121***	0.0619***	0.0787***
						(0.0271)	(0.0101)	(0.0245)
Individual commitment to civic norms						0.682***	0.290***	0.948***
						(0.106)	(0.0416)	(0.0841)
constant	2.357	2.124	0.784***	-0.803***	1.465***	-5.015***	-1.790***	
	(1.495)	(1.518)	(0.128)	(0.159)	(0.115)	(0.682)	(0.305)	
Number of observations used in the specific equation	2740	6950	8080	8060	5980	2450	2460	2440
Number of observations ever used in the whole model	8,260							

Robust standard errors in parentheses (two-sided tests)

*** p<0.01, ** p<0.05, * p<0.1

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics.
(2) Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2000-2001, 2000-2001, Version 1, U.S. Department of Education, National Center for Education Statistics. The next nine tables use the same dataset.

Table H-6 The Results for Volunteering Behaviors after Controlling for State Fixed Effects (Reading Scores from ELS:2002 as the Intermediary Variable)

VARIABLES	Read02	Math04	Extratype04	Extrahour04	Civic	Volunteer04	Voltype04	Volfreq04
User-oriented component (UOC)	0.171**	0.323**	-0.0322	0.00177	-0.00718			
	(0.0861)	(0.141)	(0.0408)	(0.0321)	(0.0126)			
UOC^2	-0.0520	-0.0966	0.00293	-0.0184	0.0101			
	(0.0556)	(0.0934)	(0.0287)	(0.0206)	(0.00891)			
Provider-oriented component (POC)	-0.000142	0.0431	-0.0380	-0.0863***	0.0178*			
	(0.0759)	(0.126)	(0.0342)	(0.0295)	(0.0107)			
POC^2	0.112**	0.129	-0.0377*	-0.0292	0.00334			
	(0.0478)	(0.0801)	(0.0217)	(0.0179)	(0.00710)			
Math score from ELS:2002						0.0323***	0.0127***	0.0173***
						(0.00458)	(0.00229)	(0.00409)
Reading score from ELS: 2002						0.0205***	0.00473**	-0.00147
						(0.00483)	(0.00221)	(0.00405)
School-centered individual socioeconomic score	3.933***	3.406***	0.517***	0.407***	0.0314***	0.242***	0.0779***	0.0902*
	(0.181)	(0.179)	(0.0389)	(0.0348)	(0.0122)	(0.0557)	(0.0222)	(0.0483)
Individual Sum of extracurricular activity types						0.334***	0.112***	0.227***
						(0.0231)	(0.00738)	(0.0142)
Individual Frequency of engaging in extracurricular activities						0.122***	0.0717***	0.0998***
						(0.0177)	(0.00806)	(0.0157)
Individual commitment to civic norms						0.416***	0.265***	0.830***
						(0.0603)	(0.0289)	(0.0550)
constant	2.281***	2.124	0.784***	-0.803***	1.465***	-4.555***	-2.158***	
	(0.792)	(1.518)	(0.128)	(0.159)	(0.115)	(0.461)	(0.233)	
Number of observations used in the specific equation	6740	6950	8080	8060	5980	5960	6020	5960
Number of observations ever used in the whole model	8270							

Robust standard errors in parentheses (two-sided tests)

*** p<0.01, ** p<0.05, * p<0.1

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics.

(2) Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2000-2001, 2000-2001, Version 1, U.S. Department of Education, National Center for Education Statistics. The next nine tables use the same dataset.

Table H-7. The Results for Volunteering Behaviors after Adjusting for Response Styles (ACT reading score as the Intermediary Variable)

VARIABLES	ACT Reading	Math04	Extratype04	Extrahour04	Civic04	Volr04	Voltype04	Volfreq04
User-oriented component (UOC)	-0.0982	0.323**	-0.0322	0.00177	-0.00718			
	(0.148)	(0.141)	(0.0408)	(0.0321)	(0.0126)			
UOC^2	0.190*	-0.0966	0.00293	-0.0184	0.0101			
	(0.104)	(0.0934)	(0.0287)	(0.0206)	(0.00891)			
Provider-oriented component (POC)	-0.0620	0.0431	-0.0380	-0.0863***	0.0178*			
	(0.163)	(0.126)	(0.0342)	(0.0295)	(0.0107)			
POC^2	-0.0467	0.129	-0.0377*	-0.0292	0.00334			
	(0.101)	(0.0801)	(0.0217)	(0.0179)	(0.00710)			
Math score from ELS:2002						0.0372***	0.0133***	0.0273***
						(0.00821)	(0.00307)	(0.00638)
Reading score from ELS: 2002						0.0249**	0.00355	-0.0106
						(0.0124)	(0.00414)	(0.00875)
School-centered individual socioeconomic score	1.791***	3.406***	0.517***	0.407***	0.0314***	0.259***	0.0529*	0.0731
	(0.185)	(0.179)	(0.0389)	(0.0348)	(0.0122)	(0.0856)	(0.0299)	(0.0713)
Individual Sum of extracurricular activity types						0.333***	0.110***	0.213***
						(0.0382)	(0.00798)	(0.0214)
Individual Frequency of engaging in extracurricular activities						0.122***	0.0624***	0.0701***
						(0.0264)	(0.0101)	(0.0240)
Individual commitment to civic norms						0.709***	0.309***	0.944***
						(0.102)	(0.0428)	(0.0848)
Acquiescence response style						0.0294**	0.00939*	0.0229**
						(0.0140)	(0.00497)	(0.0110)
Disacquiescence response style						0.0141	-0.00120	-0.00931
						(0.0180)	(0.00665)	(0.0146)
Non-contingent response style						-0.00384	0.00171	-0.00613
						(0.0153)	(0.00611)	(0.0127)
Extremely response style						-0.0653*	-0.0201	-0.112***
						(0.0341)	(0.0139)	(0.0273)
Constant	2.357	2.124	0.784***	-0.803***	1.465***	-5.031***	-1.959***	
	(1.495)	(1.518)	(0.128)	(0.159)	(0.115)	(0.675)	(0.307)	
Number of observations used in the specific equation	2740	6950	8080	8060	5980	2450	2460	2440

Number of observations ever used in the whole model	8,260

Robust standard errors in parentheses (two-sided tests)

*** p<0.01, ** p<0.05, * p<0.1

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics. (2) Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2000-2001, 2000-2001, Version 1, U.S. Department of Education, National Center for Education Statistics. The next nine tables use the same dataset.

Table H-8. The Results for Volunteering Behaviors after Adjusting for Response Styles (Reading Scores from ELS:2002 as the Intermediary Variable)

VARIABLES	Read02	Math04	extratype04k	extrahour04	civic	volunteer06	volfreq06	voltype06
User-oriented component (UOC)	0.171**	0.323**	-0.0322	0.00177	-0.00718			
	(0.0861)	(0.141)	(0.0408)	(0.0321)	(0.0126)			
UOC^2	-0.0520	-0.0966	0.00293	-0.0184	0.0101			
	(0.0556)	(0.0934)	(0.0287)	(0.0206)	(0.00891)			
Provider-oriented component (POC)	-0.000142	0.0431	-0.0380	-0.0863***	0.0178*			
	(0.0759)	(0.126)	(0.0342)	(0.0295)	(0.0107)			
POC^2	0.112**	0.129	-0.0377*	-0.0292	0.00334			
	(0.0478)	(0.0801)	(0.0217)	(0.0179)	(0.00710)			
Math score from ELS:2002						0.0329***	0.0131***	0.0167***
						(0.00457)	(0.00229)	(0.00411)
Reading score from ELS: 2002						0.0211***	0.00495**	-0.000116
						(0.00484)	(0.00221)	(0.00408)
School-centered individual socioeconomic score	3.933***	3.406***	0.517***	0.407***	0.0314***	0.235***	0.0779***	0.0829*
	(0.181)	(0.179)	(0.0389)	(0.0348)	(0.0122)	(0.0555)	(0.0220)	(0.0478)
Individual Sum of extracurricular activity types						0.320***	0.111***	0.216***
						(0.0225)	(0.00694)	(0.0140)
Individual Frequency of engaging in extracurricular activities						0.120***	0.0680***	0.0913***
						(0.0178)	(0.00809)	(0.0155)
Individual commitment to civic norms						0.420***	0.282***	0.821***
						(0.0589)	(0.0291)	(0.0549)
Acquiescence response style						0.00760	0.0104***	0.0214***
						(0.00834)	(0.00377)	(0.00699)
Disacquiescence response style						-0.00851	-0.00974*	0.00180
						(0.0103)	(0.00516)	(0.00949)
Non-contingent response style						-0.00285	0.00679	-0.00469
						(0.00860)	(0.00462)	(0.00809)
Extremely response style						-0.0692***	-0.0212**	-0.146***
						(0.0197)	(0.00982)	(0.0174)
Constant	2.281***	2.124	0.784***	-0.803***	1.465***	-3.830***	-2.247***	
	(0.792)	(1.518)	(0.128)	(0.159)	(0.115)	(0.448)	(0.226)	
Number of observations used in the specific equation	6740	6950	8080	8060	5980	5960	6020	5960

Number of observations ever used in the whole model	8,270

Robust standard errors in parentheses (two-sided tests)

*** p<0.01, ** p<0.05, * p<0.1

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics. (2) Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2000-2001, 2000-2001, Version 1, U.S. Department of Education, National Center for Education Statistics. The next nine tables use the same dataset.

Table H-9. The Results for Volunteering Behaviors in Models Distinguishing between-school and within school variation in intermediary variables (ACT reading score as the Intermediary Variable)

VARIABLES	ACT reading	Math04	Extratype04	Extrahour04	Civic04	Volr04	Voltype04	Volfreq04
User-oriented component (UOC)	-0.0982	0.323**	-0.0322	0.00177	-0.00718			
	(0.148)	(0.141)	(0.0408)	(0.0321)	(0.0126)			
UOC^2	0.190*	-0.0966	0.00293	-0.0184	0.0101			
	(0.104)	(0.0934)	(0.0287)	(0.0206)	(0.00891)			
Provider-oriented component (POC)	-0.0620	0.0431	-0.0380	-0.0863***	0.0178*			
	(0.163)	(0.126)	(0.0342)	(0.0295)	(0.0107)			
POC^2	-0.0467	0.129	-0.0377*	-0.0292	0.00334			
	(0.101)	(0.0801)	(0.0217)	(0.0179)	(0.00710)			
School-centered 2004 math score						0.0333***	0.0107***	0.0242***
						(0.00849)	(0.00329)	(0.00671)
School average 2004 math score						0.0830***	0.0297***	0.0594***
						(0.0186)	(0.00699)	(0.0156)
School-centered 2004 ACT reading						0.0234*	0.00145	-0.0130
						(0.0127)	(0.00413)	(0.00895)
School average 2004 ACT reading						0.0514*	0.0268***	0.0207
						(0.0289)	(0.0103)	(0.0230)
School-centered frequency of extracurricular activities						0.0525*	0.0349***	0.0312
						(0.0314)	(0.0109)	(0.0270)
School average frequency of extracurricular activities						0.206*	0.00273	0.00971
						(0.110)	(0.0459)	(0.0914)
School-centered types of extracurricular activities						0.258***	0.104***	0.185***
						(0.0311)	(0.00782)	(0.0200)
School average type of extracurricular activities						0.218***	0.114***	0.182***
						(0.0762)	(0.0302)	(0.0613)
School-centered endorsement of civic norms						0.688***	0.308***	0.946***
						(0.106)	(0.0416)	(0.0834)
School average endorsement of civic norms						1.291***	0.366**	1.281***
						(0.341)	(0.168)	(0.281)
Constant	2.357	2.124	0.784***	-0.803***	1.465***	-9.517***	-3.490***	
	(1.495)	(1.518)	(0.128)	(0.159)	(0.115)	(1.243)	(0.510)	

Number of observations used in the specific equation	2740	6950	8080	8060	5980	2450	2460	2440
Number of observations ever used in the whole model	8,260							

Robust standard errors in parentheses (two-sided tests)

*** p<0.01, ** p<0.05, * p<0.1

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics. (2) Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2000-2001, 2000-2001, Version 1, U.S. Department of Education, National Center for Education Statistics. The next nine tables use the same dataset.

Table H-10. The Results for Volunteering Behaviors in Models Distinguishing between-school and within school variation in intermediary variables (Reading Scores from ELS:2002 as the Intermediary Variable)

VARIABLES	Read02	Math04	Extratype04	Extrahour04	Civic04	Volr04	Voltype04	Volfreq04
User-oriented component (UOC)	0.171**	0.323**	-0.0322	0.00177	-0.00718			
	(0.0861)	(0.141)	(0.0408)	(0.0321)	(0.0126)			
UOC^2	-0.0520	-0.0966	0.00293	-0.0184	0.0101			
	(0.0556)	(0.0934)	(0.0287)	(0.0206)	(0.00891)			
Provider-oriented component (POC)	-0.000142	0.0431	-0.0380	-0.0863***	0.0178*			
	(0.0759)	(0.126)	(0.0342)	(0.0295)	(0.0107)			
POC^2	0.112**	0.129	-0.0377*	-0.0292	0.00334			
	(0.0478)	(0.0801)	(0.0217)	(0.0179)	(0.00710)			
School-centered 2004 math score						0.0329***	0.0143***	0.0208***
						(0.00466)	(0.00237)	(0.00417)
School average 2004 math score						0.0489***	0.0101	0.00867
						(0.0144)	(0.00748)	(0.0123)
School-centered 2002 ELS reading score						0.0154***	0.00264	-0.00551
						(0.00478)	(0.00224)	(0.00410)
School average 2002 ELS reading score	0.965***					0.0564***	0.0284***	0.0466***
	(0.0139)					(0.0139)	(0.00669)	(0.0113)
School-centered frequency of extracurricular activities						0.0635***	0.0459***	0.0571***
						(0.0203)	(0.00893)	(0.0177)
School average frequency of extracurricular activities						0.120	0.0100	0.0360
						(0.0852)	(0.0370)	(0.0668)
School-centered types of extracurricular activities						0.251***	0.106***	0.195***
						(0.0187)	(0.00643)	(0.0132)
School average type of extracurricular activities						0.245***	0.131***	0.131***
						(0.0562)	(0.0247)	(0.0467)
School-centered endorsement of civic norms						0.444***	0.276***	0.828***
						(0.0603)	(0.0290)	(0.0562)
School average endorsement of civic norms						0.464**	0.349***	0.949***
						(0.215)	(0.121)	(0.188)
Constant	2.281***	2.124	0.784***	-0.803***	1.465***	-7.175***	-3.628***	
	(0.792)	(1.518)	(0.128)	(0.159)	(0.115)	(0.828)	(0.384)	

Number of observations used in the specific equation	6740	6950	8080	8060	5980	5960	6020	5960
Number of observations ever used in the whole model	8270							

Robust standard errors in parentheses (two-sided tests)

*** p<0.01, ** p<0.05, * p<0.1

Data Source: (1) Educational Longitudinal Study of 2002, 2002, 2004, 2006, Version 1, U.S. Department of Education, National Center for Education Statistics. (2) Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2000-2001, 2000-2001, Version 1, U.S. Department of Education, National Center for Education Statistics. The next nine tables use the same dataset.

REFERENCES

Abedi, Jamal, Carol Lord, and Carolyn Hofstetter. 1998. "Impact of Selected Background Variables on Students' NAEP Math Performance." *CSE Technical Report* 1522 (310): 1–136.

Allison, Graham T. 2004. "Public and Private Management: Are They Fundamentally Alike in All Unimportant Respects?" In *Classics of Public Administration*, edited by Jay M Shafritz, Albert C Hyde, and Sandra J. Parkes, 396–413. Belmont,CA: Wadsworth/Thomson Learning. https://doi.org/10.2307/2212271.

Amirkhanyan, Anna A., Hyun Joon Kim, and Kristina T. Lambright. 2014. "The Performance Puzzle: Understanding the Factors Influencing Alternative Dimensions and Views of Performance." *Journal of Public Administration Research and Theory* 24 (1): 1–34. https://doi.org/10.1093/jopart/mut021.

Andersen, Simon Calmar, and Donald P. Moynihan. 2016. "Bureaucratic Investments in Expertise: Evidence from a Randomized Controlled Field Trial." *The Journal of Politics* 78 (4): 1032–44. https://doi.org/10.1086/686029.

Andersen, Simon Calmar, and Helena Skyt Nielsen. 2019. "Learning from Performance Information." *Journal of Public Administration Research and Theory* 30 (3): 1–17. https://doi.org/10.1093/jopart/muz036.

Ansell, Chris, and Alison Gash. 2008. "Collaborative Governance in Theory and Practice." *Journal of Public Administration Research and Theory* 18 (4): 543–71. https://doi.org/10.1093/jopart/mum032.

ASCD. 2015. "Comparison of the No Child Left Behind Act to Every Student Succeeds Act." https://www.ascd.org/ASCD/pdf/siteASCD/policy/ESEA_NCLB_ComparisonChart_2015.pdf.

Barton, Paul E. 2009. "National Education Standards: Getting Beneath The Surface," 1–48. papers3://publication/uuid/B6C3FBBF-EE4E-4D8B-B88E-99E67467495F.

Behn, Robert. D. 2014. *The Performance Stat Potential: A Leadership Strategy for Producing Results*. Boston: Harvard University Press.

Benaine, S. Lorenzo, and Alexander Kroll. 2020. "Explaining Effort Substitution in Performance Systems: The Role of Task Demands and Mission Orientation." *Public Management Review* 22 (6): 813–35. https://doi.org/10.1080/14719037.2019.1604794.

Berinsky, Adam J, and Gabriel S Lenz. 2011. "Education an Political Participation: Exploring the Causal Link." *Political Behavior* 33 (3): 357–73. https://doi.org/10.1007/sl.

Bevan, Gwyn, and Christopher Hood. 2006. "What Is Measured Is What Matters: Targets and Gaming in the English Public Health System." *Public Administration* 84 (3): 517–38. https://doi.org/10.1111/j.1467-9299.2006.00600.x.

Bohte, John, and Kenneth J. Meier. 2000. "Goal Displacement: Assessing the Motivation for Organizational Cheating." *Public Administration Review* 60 (2): 173–82. https://doi.org/10.1111/0033-3352.00075.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489 (7415): 295–98. https://doi.org/10.1038/nature11421.

Booher-Jennings, Jennifer. 2005. "Below the Bubble: "Educational Triage" and the Texas Accountability System." *American Educational Research Journal Summer*. Vol. 42. https://journals.sagepub.com/doi/pdf/10.3102/00028312042002231.

Boselie, Paul, Jasmijn Van Harten, and Monique Veld. 2021. "A Human Resource Management Review on Public Management and Public Administration Research: Stop Right There…before We Go Any Further…." *Public Management Review* 23 (4): 483–500. https://doi.org/10.1080/14719037.2019.1695880.

Bowman, James S. 2010. "The Success of Failure: The Paradox of Performance Pay." *Review of Public Personnel Administration* 30 (1): 70–88. https://doi.org/10.1177/0734371X09351824.

Boyne, George A., and Alex A. Chen. 2008. "Performance Targets and Public Service Improvement." *Journal of Public Administration Research and Theory* 17 (3): 455–77. https://doi.org/10.1093/jopart/mul007.

Boyne, George A., and Richard M. Walker. 2004. "Strategy Content and Public Service Organizations." *Journal of Public Administration Research and Theory* 14 (2): 231–52. https://doi.org/10.1093/jopart/muh015.

Bryk, Anthony S. 2020. *Improvement in Action : Advancing Quality in America's Schools*. Cambridge, Massachusetts: Harvard Education Press.

Bryk, Anthony S., Louis Gomez, Alicia Grunow, and LeMahieu Paul. 2015. *Learning to Improve: How America's Schools Can Get Better at Getting Better*. Boston, MA: Harvard Education Press.

Bryson, John M., Barbara C. Crosby, and Laura Bloomberg. 2014. "Public Value Governance: Moving beyond Traditional Public Administration and the New Public Management." *Public Administration Review* 74 (4): 445–56. https://doi.org/10.1111/puar.12238.

Byrne, Barbara M. 2011. *Structural Equation Modeling With Mplus*. *Structural Equation Modeling With AMOS*. New York City: Routledge. https://doi.org/10.4324/9781410600219.

Calista, Donald J. 2002. "A Critique." *Public Administration Review* 62 (3): 347–52. https://doi.org/10.1111/1540-6210.00184.

Campbell, Andrea Louise. 2012. "Policy Makes Mass Politics." *Annual Review of Political Science* 15 (1): 333–51. https://doi.org/10.1146/annurev-polisci-012610-135202.

Campbell, David E. 2006. "3 . What Is Education ' s Impact on Civic and Social Engagement ?" *Proceedings of The Copehagen Symposium*, no. March: 25–126.

———. 2008. "Voice in the Classroom: How an Open Classroom Climate Fosters Political Engagement among Adolescents." *Political Behavior* 30 (4): 437–54. https://doi.org/10.1007/s11109-008-9063-z.

———. 2009. “Civic Engagement and Education: An Empirical Test of the Sorting Model.” *American Journal of Political Science* 53 (4): 771–86.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. “Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates.” *American Economic Review* 104 (9): 2593–2632. https://doi.org/10.1257/aer.104.9.2593.

Condon, Meghan. 2015. “Voice Lessons: Rethinking the Relationship Between Education and Political Participation.” *Political Behavior* 37 (4): 819–43. https://doi.org/10.1007/s11109-015-9301-0.

Corduneanu, Roxana, Adina Dudau, and Georgios Kominis. 2020. “Crowding-in or Crowding-out: The Contribution of Self-Determination Theory to Public Service Motivation.” *Public Management Review* 22 (7): 1070–89. https://doi.org/10.1080/14719037.2020.1740303.

Couch, Arthur, and Kenneth Keniston. 1960. “Yeasayers and Naysayers: Agreeing Response Set as a Personality Variable.” *The Journal of Abnormal and Social Psychology* 60 (2): 151–74. https://doi.org/10.1037/h0040372.

Dalton, Russell J. 2008. “Citizenship Norms and the Expansion of Political Participation.” *Political Studies* 56 (1): 76–98. https://doi.org/10.1111/j.1467-9248.2007.00718.x.

Decramer, Adelien, Carine Smolders, and Alex Vanderstraeten. 2013. “Employee Performance Management Culture and System Features in Higher Education: Relationship with Employee Performance Management Satisfaction.” *International Journal of Human Resource Management* 24 (2): 352–71. https://doi.org/10.1080/09585192.2012.680602.

Dee, Thomas S. 2004. “Are There Civic Returns to Education?” *Journal of Public Economics* 88 (9–10): 1697–1720. https://doi.org/10.1016/J.JPUBECO.2003.11.002.

Dee, Thomas S., and Brian Jacob. 2011. “The Impact of No Child Left Behind on Student Achievement.” *Journal of Policy Analysis and Management* 30 (3): 418–46. https://doi.org/10.1002/pam.20586.

Dee, Thomas S., Brian Jacob, and Nathaniel L. Schwartz. 2013. “The Effects of NCLB on School Resources and Practices.” *Educational Evaluation and Policy Analysis* 35 (2): 252–79. https://doi.org/10.3102/0162373712467080.

Dee, Thomas S., and James Wyckoff. 2015. “Incentives, Selection, and Teacher Performance: Evidence from IMPACT.” *Journal of Policy Analysis and Management* 34 (2): 267–97. https://doi.org/10.1002/pam.21818.

———. 2017. “A Lasting Impact: High-Stakes Teacher Evaluations Drive Student Success in Washington, D.C.” *Education Next* 17 (4): 58–66.

DeLeon, Linda, and Robert B. Denhardt. 2000. “The Political Theory of Reinvention.” *Public Administration Review* 60 (2): 89–97. https://doi.org/10.1111/0033-3352.00068.

Denhardt, Robert B., and Janet Vinzant Denhardt. 2000. “The New Public Service: Serving Rather than Steering.” *Public Administration Review* 60 (6): 549–59. https://doi.org/10.1111/0033-3352.00117.

Dewey, John. 1916. *Democracy and Education by John Dewey Democracy and Education*. New York City: MacMillan Company. https://www.fulltextarchive.com/pdfs/Democracy-and-Education.pdf.

———. 1937. *Lectures in Educational Philosophy*. Carbondale: Southern Illinois University.

Dixit, Avinash. 2002. "Incentives and Organizations in the Public Sector: An Interpretative Review."." *Journal of Human Resources* 37 (2): 696–727.

Downs, Anthony. 1957. "An Economic Theory of Political Action in a Democracy." *Journal of Political Economy*. The University of Chicago Press. https://doi.org/10.2307/1827369.

Education Week. 2003. "Quality Counts 2003: 'If I Can't Learn from You...' Ensuring a Highly Qualified Teacher For Every Classroom." Vol. XXII. Marlon, OH.

Eisenberg, Nancy, Richard A. Fabes, Gustavo Carlo, and Mariss Karbon. 1992. "Emotional Responsivity to Others: Behavioral Correlates and Socialization Antecedents." *New Directions for Child and Adolescent Development* Spring (55): 57–73. https://doi.org/10.1002/cd.23219925506.

Favero, Nathan, Kenneth J. Meier, and Laurence J. O'Toole. 2016. "Goals, Trust, Participation, and Feedback: Linking Internal Management with Performance Outcomes." *Journal of Public Administration Research and Theory* 26 (2): 327–43. https://doi.org/10.1093/jopart/muu044.

Feldman, Amy F, and Jennifer L Matjasko. 2005. "The Role of School-Based Extracurricular Activities in Adolescent Development : A Comprehensive Review and Future Direction." *Review of Educational Research* 75 (2): 159–210.

Figlio, David N., and Lawrence S. Getzler. 2006. "Accountability, Ability and Disability: Gaming the System?" In *Improving School Accountability Chec-Ups or Choice*, edited by Timothy Gronberg and Dennis Jansen, 35–49. Somerville, MA: Emerald Group Publishing. https://doi.org/10.1016/S0278-0984(06)14002-X.

Figlio, David N., and Cecilia Elena Rouse. 2006. "Do Accountability and Voucher Threats Improve Low-Performing Schools?" *Journal of Public Economics* 90 (1–2): 239–55. https://doi.org/10.1016/J.JPUBECO.2005.08.005.

Finn, E Chester., and Diane. Ravitch. 2007. "Not By Geeks Alone." 2007. https://www.wsj.com/articles/SB118653759532491305.

Frey, Bruno S., Fabian Homberg, and Margit Osterloh. 2013. "Organizational Control Systems and Pay-for-Performance in the Public Service." *Organization Studies* 34 (7): 949–72. https://doi.org/10.1177/0170840613483655.

Fryer, Roland G. 2017. "The Production of Human Capital in Developed Countries." In *Handbook of Economic Field Experiments*, edited by Esther Duflo and Banerjee Abhijit, 2:96–174. Amsterdam, Netherland: North-Holland. https://doi.org/10.1016/bs.hefe.2016.08.006.

Gao, Jie. 2015. "Pernicious Manipulation of Performance Measures in China's Cadre Evaluation System." *China Quarterly* 223 (July 2015): 618–37. https://doi.org/10.1017/S0305741015000806.

———. 2020. “Mitigating Pernicious Gaming in Performance Management in China: Dilemmas, Strategies and Challenges.” *Public Performance and Management Review* 44 (2): 1–31. https://doi.org/10.1080/15309576.2020.1775662.

Gelman, Andrew., and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Gerrish, Ed, and Thomas Luke Spreen. 2017. “Does Benchmarking Encourage Improvement or Convergence? Evaluating North Carolina’s Fiscal Benchmarking Tool.” *Journal of Public Administration Research and Theory* 27 (4): 596–614. https://doi.org/10.1093/jopart/mux018.

Gershenson, Seth, Michael Hansen, and Constance Lindsay. 2021. *Teacher Diversity and Student Success: Why Racial Representation Matters in the Classroom*. Boston, MA: Harvard University Press. https://www.amazon.com/Teacher-Diversity-Student-Success-Representation/dp/1682535800.

Geus, Christa J.C. de, Alex Ingrams, Lars Tummers, and Sanjay K. Pandey. 2020. “Organizational Citizenship Behavior in the Public Sector: A Systematic Literature Review and Future Research Agenda.” *Public Administration Review*, 259–70. https://doi.org/10.1111/puar.13141.

Gigliotti, Philip, and Lucy C. Sorensen. 2020. “Illusory Effects of Performance Management: The Case of Contracts for Excellence in New York School Districts.” *Public Management Review*. https://doi.org/10.1080/14719037.2020.1825779.

Goodman, Sarena F, and Lesley J Turner. 2013. “The Design of Teacher Incentive Pay and Educational Outcomes: Evidence from the New York City Bonus Program.” *Source: Journal of Labor Economics* 31 (2): 409–20. https://doi.org/10.1086/668676.

Gross, Betheny, Michael Kirst, Dana Holland, and Tom Luschei. 2005. “Chapter 3 Got You Under My Spell? How Accountability Policy Is Changing and Not Changing Decision Making in High Schools.” Philadelphia. http://www.cpre.org/images/stories/cpre_pdfs/Chapter3.pdf.

Grunewald, Rob, and Anusha Nath. 2019. “Minnesota’s Education Achievement Gaps.” Washington, D.C. https://www.gao.gov/special.pubs/gao-17-776sp/resultsall.htm.

Hamilton, Laura, Brian Stecher, Julie Marsh, Jennifer McCombs, Abby Robyn, Jennifer Russell, Scott Naftel, and Heather Barney. 2007. *Standards-Based Accountability Under No Child Left Behind: Experiences of Teachers and Administrators in Three States*. *Standards-Based Accountability Under No Child Left Behind: Experiences of Teachers and Administrators in Three States*. Santa Monica, CA: RAND Corporation. https://doi.org/10.7249/mg589.

Han, Xu, and Donald Moynihan. 2021. “Does Managerial Use of Performance Information Matter to Organizational Outcomes?” *American Review of Public Administration*.

Han, Xu, and Weijie Wang. 2021. “Can Managerial Autonomy Mitigate Gaming Behaviors?” *Public Administration Review*.

Hanushek, Eric A., and Steven G. Rivkin. 2010. “Generalizations about Using Value-Added Measures of Teacher Quality.” *American Economic Review* 100 (2): 267–71. https://doi.org/10.1257/aer.100.2.267.

Hanushek, Eric A, and Margaret E Raymond. 2005. “Does School Accountability Lead to Improved Student Performance?” *Journal of Policy Analysis and Management* 24 (2): 297–327. https://doi.org/10.1002/pam.20091.

Hanushek, Eric A, and Ludger Woessmann. 2008. “The Role of Cognitive Skills in Economic Development.” *Journal of Economic Literature* 46 (3): 607–68. https://doi.org/10.1257/jel.46.3.607.

———. 2010. “The Economics OF International Differences In Educational Achievement.” 15949. http://www.nber.org/papers/w15949.

Hayes, Andrew F., and Michael Scharkow. 2013. “The Relative Trustworthiness of Inferential Tests of the Indirect Effect in Statistical Mediation Analysis.” *Psychological Science* 24 (10): 1918–27. https://doi.org/10.1177/0956797613480187.

Heinrich, Carolyn J., Patricia Burch, Annalee Good, Rudy Acosta, Huiping Cheng, Marcus Dillender, Christi Kirshbaum, Hiren Nisar, and Mary Stewart. 2014. “Low-Performing Schools Attract and Keep Academically Talented Teachers ? Evidence.” *Journal of Policy Analysis and Management* 29 (3): 471–94. https://doi.org/10.1002/pam.

Heinrich, Carolyn J, and Youseok Choi. 2007. “Performance-Based Contracting in Social Welfare Programs.” https://doi.org/10.1177/0275074006297553.

Heinrich, Carolyn J, and Gerald Marschke. 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.

Hemphill, Clara., Kim. Nauer, Helen. Zelon, Thomas. Jacobs, Alessandra. Raimondl, Sharon. McCloskey, and Rajeev. Yerneni. 2010. “Managing by the Numbers: Empowerment and Accountability in New York City’s Schools.” New York, NY.

Herd, Pamela, and Donald P. Moynihan. 2018. *Administrative Burden*. New York City: Russel Sage Foundation.

Hirsh-Pasek, Kathy, and laurence Steinberg. 2018. “Beyond the Midterms: Helping Students Overcome the Impact of No Child Left Behind.” 2018. https://www.brookings.edu/blog/education-plus-development/2018/11/21/beyond-the-midterms-helping-students-overcome-the-impact-of-no-child-left-behind/.

Hofstede, Geert. 2001. *Culture’s Consequences : Comparing Values, Behaviors, Institutions, and Organizations across Nations*. Thousand Oaks: Sage Publications.

Holm, Jakob Majlund. 2018. “Successful Problem Solvers? Managerial Performance Information Use to Improve Low Organizational Performance.” *Journal of Public Administration Research and Theory* 28 (3): 303–20. https://doi.org/10.1093/jopart/muy017.

Holmstrom, Bengt. 2016. “Bengt Holmström - Prize Lecture: Pay for Performance and Beyond.” https://doi.org/10.1257/aer.107.7.1753.

Holmstrom, Bengt, and Paul Milgrom. 1991. “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership.” *Journal of Law Economics and Organization* 7: 232–44. https://doi.org/10.1017/cbo9780511817410.018.

Hong, Sounman. 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.

Hu, Li-tze, and Peter M. Bentler. 1999. “Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives.” *Structural Equation Modeling: A Multidisciplinary Journal* 6 (1): 1–55. https://doi.org/10.1080/10705519909540118.

Huselid, Mark. 1995. “THE IMPACT OF HUMAN RESOURCE MANAGEMENT PRACTICES ON TURNOVER, PRODUCTIVITY, AND CORPORATE FINANCIAL PERFORMANCE.” *The Academy of Management Journal* 38 (3): 635–72. https://doi.org/10.1109/SPAWDA.2011.6167273.

Hvidman, Ulrik, and Simon Calmar Andersen. 2014. “Impact of Performance Management in Public and Private Organizations.” *Journal of Public Administration Research and Theory* 24 (1): 35–58. https://doi.org/10.1093/jopart/mut019.

Ingels, Steven J, Daniel J Pratt, James E Rogers, Peter H Siegel, and Ellen S Stutts. 2005. “Educational Longitudinal Study of 2002: Base-Year to First Follow-up Data File Documentation.” Washington, D.C. https://nces.ed.gov/pubs2006/2006344.pdf.

Ingels, Steven J, Daniel J Pratt, David Wilson, Laura J Burns, Douglas Currivan, James E Rogers, and Sherry Hubbard-Bednasz. 2007. “Education Longitudinal Study of 2002 (ELS:2002): Base-Year to Second Follow-up Data File Documentation.” Washington, D.C. https://pdfs.semanticscholar.org/bb67/333b39805225c447a960acb98547c6b64877.pdf.

Jacob, Brian A. 2005. “Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools.” *Journal of Public Economics* 89 (5–6): 761–96. https://doi.org/10.1016/J.JPUBECO.2004.08.004.

Jacob, Brian A., and Steven D. Levitt. 2003. “Rotten Apples: An Investigation of The Prevalence and Predictors of Teacher Cheating.” *The Quarterly Journal of Economics*, no. August: 843–77.

Jacobsen, Christian Bøtcher, and Lotte Bøgh Andersen. 2014. “Performance Management in the Public Sector: Does It Decrease or Increase Innovation and Performance?” *International Journal of Public Administration* 37 (14): 1011–23. https://doi.org/10.1080/01900692.2014.928317.

Jakobsen, Mads L.F., and Peter B. Mortensen. 2016. “Rules and the Doctrine of Performance Management.” *Public Administration Review* 76 (2): 302–12. https://doi.org/10.1111/puar.12409.

James, Oliver, Asmus Leth Olsen, Donald Moynihan, and Gregg G Van Ryzin. 2020. *Behavioral Public Performance: How People Make Sense of Government Metrics*. New York, NY: Cambridge University Press.

Jilke, Sebastian, and Lars Tummers. 2018. “Which Clients Are Deserving of Help? A Theoretical Model and Experimental Test.” *Journal of Public Administration Research and Theory* 28 (2): 226–38.

Johnson, Timothy, Patrick Kulesa, Young IK Cho, and Sharon Shavitt. 2005. “The Relation

Between Culture and Response Styles Evidence from 19 Countries." *Culture and Response Styles Journal of Cross-Cultural Psychology* 36 (2): 264–77. https://doi.org/10.1177/0022022104272905.

Joyce, Philip G. 2007. "Performance and Budgeting Under the Separation of Powers." In *Performance Budgeting Linking Funding and Results*, edited by Marc Robinson, 442–61. New York: Palgrave Macmillan.

Kam, Cindy D., and Carl L. Palmer. 2008. "Reconsidering the Effects of Education on Political Participation." *Journal of Politics* 70 (3): 612–31. https://doi.org/10.1017/S0022381608080651.

Kelman, Steven, and John N. Friedman. 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): 917–46. https://doi.org/10.1093/jopart/mun028.

Kenny, David. 2018. "Mediation." 2018. http://davidakenny.net/cm/mediate.htm.

Klandermans, Bert. 1984. "Mobilization and Participation: Social-Psychological Expansisons of Resource Mobilization Theory." *American Sociological Review*. American Sociological Association. https://doi.org/10.2307/2095417.

Kleine, Andrew. 2018. *City on the Line : How Baltimore Transformed Its Budget to Beat the Great Recession and Deliver Outcomes*. 1st ed. Lanham, Maryland: Rowman & Littlefield.

Koning, Pierre, and Carolyn J. Heinrich. 2013. "Cream-Skimming, Parking and Other Intended and Unintended Effects of High-Powered, Performance-Based Contracts." *Journal of Policy Analysis and Management* 32 (3): 461–83. https://doi.org/10.1002/PAM.21695.

Krieg, John M. 2008. "Are Students Left Behind? The Distributional Effects of the No Child Left Behind Act." *Education Finance and Policy* 3 (2): 250–81. https://doi.org/10.1162/edfp.2008.3.2.250.

Kroll, Alexander. 2017. "Can Performance Management Foster Social Equity? Stakeholder Power, Protective Institutions, and Minority Representation." *Public Administration* 95 (1): 22–38. https://doi.org/10.1111/padm.12289.

Kroll, Alexander, and Donald P. Moynihan. 2015. "Does Training Matter? Evidence from Performance Management Reforms." *Public Administration Review* 75 (3): 411–20. https://doi.org/10.1111/puar.12331.

Kroll, Alexander, Milena I. Neshkova, and Sanjay K. Pandey. 2019. "Spillover Effects From Customer to Citizen Orientation: How Performance Management Reforms Can Foster Public Participation." *Administration and Society* 51 (8): 1227–53. https://doi.org/10.1177/0095399716687341.

Krueger, Alan B. 2003. "Economic Considerations and Class Size Author (s): Alan B . Krueger Source : The Economic Journal , Vol . 113 , No . 485 , Features (Feb ., 2003), Pp . F34-F63 Published by : Wiley on Behalf of the Royal Economic Society Stable URL : Https://Www.Jstor." *The Economic Journal* 113 (485): 34–63.

Kumlin, Staffan, and Bo Rothstein. 2005. “Making and Breaking Social Capital.” *Comparative Political Studies* 38 (4): 339–65. https://doi.org/10.1177/0010414004273203.

Ladd, Helen F., and Douglas L. Lauen. 2010. “Status versus Growth: The Distributional Effects of School Accountability Policies.” *Journal of Policy Analysis and Management* 29 (3): 426–50. https://doi.org/10.1002/pam.20504.

Lanyon, Richard I., and Leonard David Goodstein. 1982. *Personality Assessment*. New York City: Wiley.

Lawford, Heather, Michael W. Pratt, Bruce Hunsberger, and S. Mark Pancer. 2005. “Adolescent Generativity: A Longitudinal Study of Two Possible Contexts for Learning Concern for Future Generations.” *Journal of Research on Adolescence* 15 (3): 261–73. https://doi.org/10.1111/j.1532-7795.2005.00096.x.

Lee, Jaekyung, and Todd Reeves. 2012. “Revisiting the Impact of NCLB High-Stakes School Accountability, Capacity, and Resources: State NAEP 1990-2009 Reading and Math Achievement Gaps and Trends.” *Educational Evaluation and Policy Analysis* 34 (2): 209–31. https://doi.org/10.3102/0162373711431604.

Lee, Valerie E., and Anthony S. Bryk. 1989. “A Multilevel Model of the Social Distribution of High School Achievement.” *Sociology of Education* 62 (3): 172. https://doi.org/10.2307/2112866.

Lewis, Gregory B., and Rahul Pathak. 2014. “The State and Local Government Workforce: Trends and Patterns in Georgia, the South, and the United States.” http://cslf.gsu.edu/files/2014/10/CSLF2_FIN.pdf.

Lipsky, Michael. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Service*. New York, NY: Russell Sage Foundation.

Locke, Edwin A, and Gary P Latham. 2002. “Building a Practically Useful Theory of Goal Setting and Task Motivation A 35-Year Odyssey.” *American Psychologist* 57 (9): 705–17. https://doi.org/10.1037/0003-066X.57.9.705.

Ludwig, Jens, Jeffrey R. Kling, and Sendhil Mullainathan. 2017. “Social Policy: Mechanism Experiments and Policy Evaluations.” In *Handbook on Field Experiment*, edited by Esther Duflo and Banerjee Abhijit, 25:391–426. Amsterdam: North Holland. https://doi.org/10.1257/jep.25.3.17.

MacKinnon, David P., Matthew S. Fritz, Jason Williams, and Chondra M. Lockwood. 2007. “Distribution of the Product Confidence Limits for the Indirect Effect: Program PRODCLIN.” *Behavior Research Methods* 39 (3): 384–89. https://doi.org/10.3758/BF03193007.

Massell, Diane, Margaret Goertz, Gayle Christensen, and Matthew Goldwasser. 2005. “Chapter 2 The Press From Above, the Pull From Below: High School Responses to External Accountability.” Philadelphia. http://ww.w.learning-connections.org/images/stories/cpre_pdfs/Chapter2.pdf.

Mayer, Gerald. 2014. “Selected Characteristics of Private and Public Sector Workers.” Washington, D.C. www.crs.gov.

Mccaffrey, Daniel F., Beth Ann Griffin, Daniel Almirall, Mary Ellen Slaughter, Rajeev Ramchand, and Lane F. Burgette. 2013. "A Tutorial on Propensity Score Estimation for Multiple Treatments Using Generalized Boosted Models." *Statistics in Medicine* 32 (19): 3388–3414. https://doi.org/10.1002/sim.5753.

McFarland, Daniel A., and Reuben J. Thomas. 2006. "Bowling Young: How Youth Voluntary Associations Influence Adult Political Participation." *American Sociological Review* 71 (3): 401–25. https://doi.org/10.1177/000312240607100303.

McLaughlin, Don, Victor Bandeira De Mello, Charles Blankenship, Kassandra. Chaney, Phil. Esra, Hiro. Hkawa, Daniela. Rojas, Paul. William, Michael. Wolman, and Tasilma. Rahman. 2008. "Comparison Between NAEP and State Mathematics Assessment Results: 2003." Vol. 1. Washington, D.C.

Mehta, Jal. 2013. "How Paradigms Create Politics." *American Educational Research Journal* 50 (2): 285–324. https://doi.org/10.3102/0002831212471417.

Meier, Kenneth J., Nathan Favero, and Ling Zhu. 2015. "Performance Gaps and Managerial Decisions: A Bayesian Decision Theory of Managerial Action." *Journal of Public Administration Research and Theory* 25 (4): 1221–46. https://doi.org/10.1093/jopart/muu054.

Meier, Kenneth J., Laurence J. O'Toole Jr, George A. Boyne, and Richard M. Walker. 2006. "Strategic Management and the Performance of Public Organizations: Testing Venerable Ideas against Recent Theories." *Journal of Public Administration Research and Theory* 17 (3): 357–77. https://doi.org/10.1093/jopart/mul017.

Meier, Kenneth J., and Laurence J. O'Toole. 2008. "Management Theory and Occam's Razor." *Administration & Society* 39 (8): 931–58. https://doi.org/10.1177/0095399707309355.

———. 2013. "Subjective Organizational Performance and Measurement Error: Common Source Bias and Spurious Relationships." *Journal of Public Administration Research and Theory* 23 (2): 429–56. https://doi.org/10.1093/jopart/mus057.

Meier, Kenneth J., Laurence J. O'Toole, and Alisa Hicklin. 2010. "I've Seen Fire and I've Seen Rain." *Administration & Society* 41 (8): 979–1003. https://doi.org/10.1177/0095399709349027.

Mettler, Suzanne, and Joe Soss. 2004. "The Consequences of Public Policy for Democratic Citizenship: Bridging Policy Studies and Mass Politics." *Perspectives on Politics* 2 (1): 55–73. https://doi.org/10.1017/S1537592704000623.

Milligan, Kevin, Enrico Moretti, and Philip Oreopoulos. 2004. "Does Education Improve Citizenship? Evidence from the United States and the United Kingdom." *Journal of Public Economics* 88 (9–10): 1667–95. https://doi.org/10.1016/J.JPUBECO.2003.10.005.

Moe, Terry M. 2009. "Collective Bargaining and The Performance of the Public Schools." *American Political Science Journal* 53 (1): 156–74.

Moynihan, Donald. P., Sergio. Fernandez, Soonhee. Kim, Kelly. M. LeRoux, Suzanne. J. Piotrowski, Bradley. E. Wright, and Kaifeng. Yang. 2011. "Performance Regimes Amidst Governance Complexity." *Journal of Public Administration Research and Theory* 21 (Supplement 1): i141–55. https://doi.org/10.1093/jopart/muq059.

Moynihan, Donald P. 2008. *The Dynamics of Performance Management : Constructing Information and Reform*. Washington, D.C: Georgetown University Press.

———. 2010. "A Workforce of Cynics? The Effects of Contemporary Reforms on Public Service Motivation." *International Public Management Journal* 13 (1): 24–34. https://doi.org/10.1080/10967490903547167.

———. 2015. "Avoiding Performance Perversity: When Incentives Lead to Deteriorating Services." 2015. http://www.performwell.org/index.php/webinars/321-avoiding-performance-perversity-when-incentives-lead-to-deteriorating-services.

Moynihan, Donald P., and Ivor Beazley. 2016. "Toward Next-Generation Performance Budgeting Lessons from the Experiences of Seven Reforming Countries Public Sector Governance." Washington, DC. http://documents.worldbank.org/curated/en/356081478497402740/pdf/109808-PUB-Box396311B-PUBLIC-DOCDATE-11-1-16.pdf.

Moynihan, Donald P., and Noel Landuyt. 2009. "How Do Public Organizations Learn? Bridging Cultural and Structural Perspectives." *Public Administration Review* 69 (6): 1097–1105. https://doi.org/10.1111/j.1540-6210.2009.02067.x.

Moynihan, Donald P., and Joe Soss. 2014. "Policy Feedback and the Politics of Administration." *Public Administration Review* 74 (3): 320–32. https://doi.org/10.1111/puar.12200.

Mulaik, Stanley A. 2010. *Foundations of Factor Analysis*. 2nd Editio. Boca Raton, Florida: CRC Press.

Muralidharan, K. 2017. *Field Experiments in Education in Developing Countries*. *Handbook of Economic Field Experiments*. Vol. 2. Elsevier Ltd. https://doi.org/10.1016/bs.hefe.2016.09.004.

Murnane, Richard J., and John B. Willett. 2011. *Methods Matter : Improving Causal Inference in Educational and Social Science Research*. New York : Oxford University Press.

Musick, Marc A., and John Wilson. 2008. *Volunteers : A Social Profile*. Bloomington, IN: Indiana University Press. https://books.google.com/books/about/Volunteers.html?id=u8Tabf5HcRcC.

Musick, Marc A. 2008. "The Importance of Studying Volunteering." In *Volunteers,A Social Profile*.

Musso, Juliet A., Matthew M. Young, and Michael Thom. 2019. "Volunteerism as Co-Production in Public Service Management: Application to Public Safety in California." *Public Management Review* 21 (4): 473–94. https://doi.org/10.1080/14719037.2018.1487574.

Mutz, Diana C. 2002. "The Consequences of Cross-Cutting Networks for Political Participation." *American Journal of Political Science* 46 (4): 838. https://doi.org/10.2307/3088437.

Naemi, Bobby D, Daniel J Beal, and Stephanie C Payne. 1992. "Personality Predictors of Extreme Response Style." *Journal of Personality* 77: 1. https://doi.org/10.1111/j.1467-6494.2008.00545.x.

National Center for Educational Statistics. 2019. “Mathematics Achievement Levels by Grade.” 2019. https://nces.ed.gov/nationsreportcard/mathematics/achieve.aspx.

Neal, Derek. 2011. “The Design of Performance Pay in Education.” In *Handbook of the Economics of Education*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 4:495–550. Elsevier B.V. https://doi.org/10.1016/B978-0-444-53444-6.00006-7.

Neal, Derek, and Diane Whitmore Schanzenbach. 2010. “Left Behind By Design: Proficiency Counts and Test-Based Accountability.” *The Review of Economics and Statistics,* 92 (2): 263–83. https://www.mitpressjournals.org/doi/pdf/10.1162/rest.2010.12318.

Nie, Norman H., Jane. Junn, and Kenneth. Stehlik-Barry. 1996. *Education and Democratic Citizenship in America*. University of Chicago Press.

Nielsen, Poul A. 2014a. “Learning from Performance Feedback: Performance Information, Aspiration Levels, and Managerial Priorities.” *Public Administration* 92 (1): 142–60. https://doi.org/10.1111/padm.12050.

———. 2014b. “Performance Management, Managerial Authority, and Public Service Performance.” *Journal of Public Administration Research and Theory* 24 (2): 431–58. https://doi.org/10.1093/jopart/mut025.

Nielsen, Poul A., and Donald P. Moynihan. 2017. “How Do Politicians Attribute Bureaucratic Responsibility for Performance? Negativity Bias and Interest Group Advocacy.” *Journal of Public Administration Research and Theory* 27 (2): muw060. https://doi.org/10.1093/jopart/muw060.

O’Toole, Laurence J, and Kenneth J. Meier. 2004. “Desperately Seeking Selznick: Cooptation and the Dark Side of Public Management in Networks.” *Public Administration Review* 64 (6): 681–93. https://doi.org/10.1111/j.1540-6210.2004.00415.x.

———. 2011. *Public Management Organizations, Governance, Performance*. Cambridge, UK: Cambridge University Press.

O’Toole, Laurence J, and Kenneth J Meier. 2015. “Public Management, Context, and Performance: In Quest of a More General Theory.” *Journal of Public Administration Research and Theory* 25 (1): 237–56. https://doi.org/10.1093/jopart/muu011.

OECD. 2017. “OECD Guidelines on Measuring Trust.” *OECD Guidelines on Measuring Trust*. Paris. https://doi.org/10.1787/9789264278219-en.

Olsen, Asmus Leth. 2017. “Compared to What? How Social and Historical Reference Points Affect Citizens’ Performance Evaluations.” *Journal of Public Administration Research and Theory* 27 (4): 562–80. https://doi.org/10.1093/jopart/mux023.

Olson, Mancur. 1971. *The Logic of Collective Action; Public Goods and the Theory of Groups.* Harvard University Press.

Osborne, David. 2017. *Reinventing America’s Schools: Creating a 21st Century Education System*. New York: Bloomsbury. https://www.amazon.com/dp/B0731CFQBG/ref=dp-kindle-redirect?_encoding=UTF8&btkr=1.

Osborne, David, and Ted Gaebler. 1992. *Reinventing Government -How the Enrepreneurial Spirit Is Transforming the Public Sector*. New York City: Addison-Wesley Publishing Company, Inc.

Osborne, Stephen P., Zoe Radnor, and Kirsty Strokosch. 2016. "Co-Production and the Co-Creation of Value in Public Services: A Suitable Case for Treatment?" *Public Management Review* 18 (5): 639–53. https://doi.org/10.1080/14719037.2015.1111927.

Ouchi, William G., and Lydia G. Segal. 2003. *Making Schools Work A Revolutionary Plan to Get Your Children the Education They Need*. New York: Simon & Schuster.

Perry, James L., Trent A. Engbers, and So Yun Jun. 2009. "Back to the Future? Performance-Related Pay, Empirical Research, and the Perils of Persistence." *Public Administration Review* 69 (1): 39–51. https://doi.org/10.1111/j.1540-6210.2008.01939_2.x.

Podgursky, Michael J., and Matthew G. Springer. 2007. "Teacher Performance Pay: A Review." *Journal of Policy Analysis and Management* 26 (4): 909–50. https://doi.org/10.1002/pam.20292.

Podsakoff, Philip M., Scott B. MacKenzie, and Nathan P. Podsakoff. 2012. "Sources of Method Bias in Social Science Research and Recommendations on How to Control It." *Annual Review of Psychology* 63 (1): 539–69. https://doi.org/10.1146/annurev-psych-120710-100452.

Poister, Theodore. H, Obed. Q Pasha, and Lauren Hamilton Edwards. 2013. "Does Performance Management Lead to Better Outcomes? Evidence from the U.S. Public Transit Industry." *Public Administration Review* 73 (4): 625–36. https://doi.org/10.1111/puar.12076. Does.

Pollitt, Christopher, and Geert Bouckaert. 2017. *Public Management Reform*. Oxford, England: Oxford University Press. https://doi.org/10.1017/CBO9781107415324.004.

Primo, David M., Matthew L. Jacobsmeier, and Jeffrey Milyo. 2007. "Estimating the Impact of State Policies and Institutions with Mixed-Level Data." *State Politics & Policy Quarterly* 7 (4): 446–59. https://doi.org/10.1177/153244000700700405.

Radin, Beryl A. 2006. *Challenging the Performance Movement : Accountability, Complexity, and Democratic Values*. Georgetown University Press.

Raffel, Jeffrey A. 2007. "Why Has Public Administration Ignored Public Education, and Does It Matter?" *Public Administration Review* 67 (1): 135–51. https://doi.org/10.1111/j.1540-6210.2006.00703.x.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models : Applications and Data Analysis Methods*. Thousand Oak, CA: Sage Publications.

Rawls, John. 2001. *Justice as Fairness : A Restatement*. Edited by Erin Kelly. Cambridge, MA: Harvard University Press. http://www.hup.harvard.edu/catalog.php?isbn=9780674005112.

Reback, Randall. 2008. "Teaching to the Rating: School Accountability and the Distribution of Student Achievement." *Journal of Public Economics* 92 (5–6): 1394–1415. https://doi.org/10.1016/J.JPUBECO.2007.05.003.

Ridgeway, Greg, Dan McCaffrey, Andrew Morral, Lane Burgette, and Beth Ann Griffin. 2017. "Toolkit for Weighting and Analysis of Nonequivalent Groups: A Tutorial for the Twang Package." http://users.nber.org/~rdehejia/.

Rockefeller, Steven C. 1993. "Comment on Charles Taylor's Essay." In *Multiculturalism: Examining the Politics of Recognition*, edited by Amy Gutmann, 89–91.

Rockoff, By Jonah E, Douglas O Staiger, Thomas J Kane, and Eric S Taylor. 2012. "Information and Employee Evaluation : Evidence from a Randomized Intervention in Public Schools." *The American Economic Review* 102 (7): 3184–3213.

Rosenstone, Steven J., and John Mark. Hansen. 1993. *Mobilization, Participation, and Democracy in America*. Macmillan Pub. Co. https://books.google.com/books/about/Mobilization_Participation_and_Democracy.html?id=QyuHAAAAMAAJ.

Rosenthal, Saul, Candice Feiring, and Michael Lewis. 1998. "Political Volunteering From Late Adolescence to Young Adulthood." *Journal of Social Issues* 54 (3): 477–93. https://doi.org/10.1111/j.1540-4560.1998.tb01231.x.

Rothstein, Jesse. 2016. "Can Value-Added Models Identify Teachers' Impacts? - Institute for Research on Labor and Employment." December 21, 2016. https://irle.berkeley.edu/can-value-added-models-identify-teachers-impacts/.

Rothstein, Richard., Rebecca. Jacobsen, and Tamara Wilder. 2008. *Grading Education : Getting Accountability Right*. Washington, D.C & New York City: Economic Policy Institute & Teachers College Press.

RTI. 2002. "Student Questionair Base Year 10th Grade." Research Triangle Park, NC. https://nces.ed.gov/surveys/els2002/pdf/StudentQ_baseyear.pdf.

———. 2004. "Student Questionair First Follow-Up." Research Triangle Park, NC. https://nces.ed.gov/surveys/els2002/pdf/FinalStudent_followup1.pdf.

Schuster, Christian, Jan Hinrik Meyer-Sahling, and Kim Sass Mikkelsen. 2020. "(Un)Principled Principals, (Un)Principled Agents: The Differential Effects of Managerial Civil Service Reforms on Corruption in Developing and OECD Countries." *Governance*, no. August: 1–20. https://doi.org/10.1111/gove.12461.

Skocpol, Theda. 1992. *Protecting Soldiers and Mothers : The Political Origins of Social Policy in the United States*. Boston: Belknap Press.

Snijders, Tom, and Roel Bosker. 2012. *Multilevel Analysis : An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: SAGE Publications Ltd.

Snyder, Jeffrey W, Andrew Saultz, and Rebecca. Jacobsen. 2017. "Antipolitics and the Hindrance of Performance Management in Education." *Journal of Public Administration Research and Theory*, October, 1–7. https://doi.org/10.1093/jopart/mux033.

Sondheimer, Rachel Milstein, and Donald P. Green. 2010. "Using Experiments to Estimate the Effects of Education on Voter Turnout." *American Journal of Political Science* 54 (1): 174–89. https://doi.org/10.1111/j.1540-5907.2009.00425.x.

Soss, Joe., and Sanford F. Schram. 2007. "A Public Transformed? Welfare Reform as Policy Feedback." *American Political Science Review* 101 (1): 111–27. https://doi.org/10.1017/S0003055407070049.

Soss, Joe. 1999. "Lessons of Welfare: Policy Design, Political Learning, and Political Action." *American Political Science Review* 93 (02): 363–80. https://doi.org/10.2307/2585401.

———. 2002. *Unwanted Claims : The Politics of Participation in the U.S. Welfare System*. Ann Arbor: University of Michigan Press.

Soss, Joe, Richard Fording, and Sanford F. Schram. 2011. "The Organization of Discipline: From Performance Management to Perversity and Punishment." *Journal of Public Administration Research and Theory* 21 (SUPPL. 2): i203–32. https://doi.org/10.1093/jopart/muq095.

Spreen, Thomas Luke, Whitney Afonso, and Ed Gerrish. 2020. "Can Employee Training Influence Local Fiscal Outcomes?" *American Review of Public Administration* 50: 1–14. https://doi.org/10.1177/0275074020911717.

Springer, Matthew G. 2008. "The Influence of an NCLB Accountability Plan on the Distribution of Student Test Score Gains." *Economics of Education Review* 27 (5): 556–63. https://doi.org/10.1016/j.econedurev.2007.06.004.

Springer, Matthew G, Dale Ballou, Laura S Hamilton, Vi-nhuan Le, J R Lockwood, Daniel F Mccaffrey, Matthew Pepper, and Brian M Stecher. 2010. "Teacher Pay for Performance Experimental Evidence from The." Nashville, Tennessee.

Springer, Matthew G, John F Pane, Vi-Nhuan Le, Daniel F Mccaffrey, Susan Freeman Burns, Laura S Hamilton, and Brian Stecher. 2012. "Team Pay for Performance: Experimental Evidence From the Round Rock Pilot Project on Team Incentives." *Source: Educational Evaluation and Policy Analysis* 34 (4): 367–90. https://doi.org/10.3102/0162373712439094.

Taylor, Eric S, and John H Tyler. 2012. "The Effect of Evaluation on Teacher Performance." *American Economic Review* 102 (7): 3628–51. https://doi.org/10.1257/aer.102.7.3628.

Tenn, Steven. 2007. "The Effect of Education on Voter Turnout." *Political Analysis* 15 (4): 446–64. https://doi.org/10.1093/pan/mpm012.

Terman, Jessica N., and Kaifeng Yang. 2016. "Reconsidering Gaming in an Accountability Relationship: The Case of Minority Purchasing in Florida." *Public Performance and Management Review* 40 (2): 281–309. https://doi.org/10.1080/15309576.2016.1177560.

The Urban Institute-Brookings Institute Tax Policy Center. 2017. "State and Local General Expenditures, Per Capital." 2017. https://www.taxpolicycenter.org/statistics/state-and-local-general-expenditures-capita.

Tofighi, Davood, and David P. MacKinnon. 2011. "RMediation: An R Package for Mediation Analysis Confidence Intervals." *Behavior Research Methods* 43 (3): 692–700. https://doi.org/10.3758/s13428-011-0076-x.

Tummers, Lars L.G., Victor Bekkers, Evelien Vink, and Michael Musheno. 2015. "Coping during Public Service Delivery: A Conceptualization and Systematic Review of the Literature." *Journal of Public Administration Research and Theory* 25 (4): 1099–1126.

https://doi.org/10.1093/jopart/muu056.

U.S. Department of Education. 2005. "No Child Left Behind: A Roadmap for State Implementation." Vol. 2006. http://www2.ed.gov/admins/lead/account/roadmap/roadmap.pdf.

———. 2017. "Every Student Succeeds Act (ESSA) | U.S. Department of Education." 2017. https://www.ed.gov/essa?src=rn.

Verba, Sidney., Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and Equality : Civic Voluntarism in American Politics*. Cambridge, MA: Harvard University Press.

Vroom, Victor H. 1964. *Work and Motivation*. New York City: John Wiley &sons.

Waeyenberg, Thomas Van, Adelien Decramer, Sebastian Desmidt, and Mieke Audenaert. 2017. "The Relationship between Employee Performance Management and Civil Servants' Turnover Intentions: A Test of the Mediating Roles of System Satisfaction and Affective Commitment." *Public Management Review* 19 (6): 747–64. https://doi.org/10.1080/14719037.2016.1209230.

Wang, Weijie, and Ryan Yeung. 2018. "Testing the Effectiveness of 'Managing for Results': Evidence from an Education Policy Innovation in New York City." *Journal of Public Administration Research and Theory* 34: 1–17. https://doi.org/10.1093/jopart/muy043.

Watkins, David, and Steven Cheung. 1995. "Culture, Gender, and Response Bias." *Journal of Cross-Cultural Psychology* 26 (5): 490–504. https://doi.org/10.1177/0022022195265003.

Weisberg, Daniel, Susan Sexton, Jennifer Mulhern, and David Keelin. 2009. "The Widget Effect." *The New Teacher Project*. New York, NY. https://tntp.org/assets/documents/TheWidgetEffect_2nd_ed.pdf.

Wichowsky, Amber, and Donald P. Moynihan. 2008. "Measuring How Administration Shapes Citizenship: A Policy Feedback Perspective on Performance Management." *Public Administration Review* 68 (5): 908–20. https://doi.org/10.1111/j.1540-6210.2008.00931.x.

Wilson, John. 2000. "Volunteering." *Annual Review of Sociology* 26: 215–40.

Zhang, Zhen, Michael J. Zyphur, and Kristopher J. Preacher. 2009. "Testing Multilevel Mediation Using Hierarchical Linear Models: Problems and Solutions." *Organizational Research Methods* 12 (4): 695–719. https://doi.org/10.1177/1094428108327450.