

5-18-2020

Dysfunctional Uses of Performance Information in Public Organizations: Understanding the Determinants of Gaming Behavior

Shawn Benaine
sben004@fiu.edu

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>



Part of the [Other Public Affairs, Public Policy and Public Administration Commons](#), [Public Administration Commons](#), and the [Public Affairs Commons](#)

Recommended Citation

Benaine, Shawn, "Dysfunctional Uses of Performance Information in Public Organizations: Understanding the Determinants of Gaming Behavior" (2020). *FIU Electronic Theses and Dissertations*. 4527.
<https://digitalcommons.fiu.edu/etd/4527>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

DYSFUNCTIONAL USES OF PERFORMANCE INFORMATION IN PUBLIC
ORGANIZATIONS: UNDERSTANDING THE DETERMINANTS OF GAMING
BEHAVIOR

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PUBLIC AFFAIRS

by

Shawn Lorenzo Benaine

2020

To: Dean John F. Stack, Jr.
Steven J. Green School of International and Public Affairs

This dissertation, written by Shawn Lorenzo Benaine and entitled *Dysfunctional Uses of Performance Information in Public Organizations: Understanding the Determinants of Gaming Behavior*, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Can Chen

Milena Neshkova

Nathan Hiller

Alexander Kroll, Major Professor

Date of Defense: May 18, 2020

The dissertation of Shawn Lorenzo Benaine is approved.

Dean John F. Stack, Jr.
Steven J. Green School of International and Public Affairs

Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2020

© Copyright 2020 by Shawn Lorenzo Benaine

All rights reserved.

DEDICATION

To my parents, Donald and Sandra Benaine, for their love and support. And for instilling me an admiration and a desire for knowledge

ACKNOWLEDGMENTS

I acknowledge God above all for allowing me the opportunity and the blessing to pursue my research and complete my dissertation. I also thank my parents, who I have dedicated this dissertation to, and all members of my family and close friends for their continuous support. In addition, I thank my fellow students in my graduate cohort, as well as others in my university department and beyond. I thank my dissertation chair, Alex, for his guidance and his expert knowledge, as well as my committee members. I thank the hundreds of participants (school leaders and Amazon M-Turkers) that took part in my research and the various organizations, such as United Teachers of Dade and the School Board of Miami-Dade County, for their assistance during my data collection. I also thank Emerald Publishing Limited for granting me permission to reprint my first essay. I published my first essay of my dissertation in the International Journal of Public Sector Management, a peer review journal and owned by Emerald Publishing. I have included the written permission on the first page of the reprint at their request. There are countless others who I cannot name specifically because of space, but I have thanked them personally. A dissertation is the work and masterpiece of an individual, but the product of a community and for the benefit of society at large. Without the support of a community of family, friends, and professional peers and mentors, as well as research participants, a dissertation would be impossible to complete. And so, I conclude my acknowledgements by thanking all those in this familial community.

ABSTRACT OF THE DISSERTATION
DYSFUNCTIONAL USES OF PERFORMANCE INFORMATION IN PUBLIC
ORGANIZATIONS: UNDERSTANDING THE DETERMINANTS OF GAMING
BEHAVIOR

by

Shawn Lorenzo Benaine

Florida International University, 2020

Miami, Florida

Professor Alexander Kroll, Major Professor

The use of performance information in decision-making is central to the practice of public management and is a common feature of modern governance (Kroll, 2015; Moynihan, 2008). The purposeful use of performance information brings about many benefits for public organizations, such as enhanced democratic accountability and responsiveness, as well as improved managerial decision-making in the public interest. However, there are dysfunctional uses of performance systems, such as performance gaming. This is a behavior where organizations or individuals manipulate data or effort in a way that undermines organizational goals. The dissertation seeks to understand what factors mitigate performance gaming. Prior research in public administration has focused heavily on system factors as the drivers of performance gaming. However, other social science disciplines have examined behavioral characteristics. This dissertation draws on these disciplines to make the argument that behavioral traits, particularly other-oriented constructs (i.e., mission orientation and prosocial impact), have the potential to mitigate performance gaming. The document consists of three essays: 1) a systematic literature

review, 2) an observational study, using a five-year panel data set of high schools, and 3) a two-part, randomized, experimental study, using a sample of general population participants and a sample of public managers (school leaders). The findings show that other-oriented constructs, such as mission orientation and prosocial impact, do mitigate performance gaming. Specifically, mission orientation reduces the effect of performance gaming when organizations have demanding tasks. Prosocial impact lessens performance gaming when public managers give more attention to performance data or when they are less experienced. This dissertation contributes to theory by highlighting the importance of other-oriented constructs for performance gaming behavior within public organizations. For example, performance information use among citizens and decision-makers can be different in many ways, such as performance gaming is affected by the role of beneficiaries or the time spent. It also contributes to public management practice by providing recommendations for decision-making based on the influence of other-oriented constructs. The findings have broader implications for the welfare of the general public, including directions for future research on performance gaming's effect on historically disadvantaged groups.

TABLE OF CONTENTS

CHAPTER	subheadings	PAGE
I. INTRODUCTION.....		1
1.1 Purpose of the Study		1
1.1.1 Importance of the topic.....		1
1.1.2 Definition of Central Concepts.....		3
1.1.3 Research question: What are the determinants of performance gaming?.....		4
1.2 State of the Literature and Research Gaps		5
1.2.1 Principal-Agent Theory and Performance Gaming.....		6
1.2.2 The Theory of Mission Orientation		8
1.2.3 Conclusion of Theoretical Foundation		10
1.3 Argument of the Dissertation		10
1.3.1 The Behavioral Perspective for Understanding Performance Gaming		11
1.3.2 Mission-Orientation in Public Organizations		12
1.3.3 Prosocial Impact and Visualizations of Performance Data		12
1.3.4 Summary of the Main Argument.....		13
1.4 Approach of the Dissertation.....		15
1.5 Explanation of the Three Essays		18
II. ESSAY 1: PERFORMANCE GAMING: A SYSTEMATIC REVIEW OF THE LITERATURE IN PUBLIC ADMINISTRATION.....		22
2.1 Introduction.....		23
2.2 Methods.....		24
2.2.1 Systematic literature review		25
2.2.2 PA Research: Systems and Organizations.....		42
2.2.3 Behavioral Insights from Other Disciplines		43
2.3 Analysis of empirical studies		45
2.3.1 System Factors.....		45
2.3.2 Organizational Characteristics.....		46
2.3.3 Individual characteristic		47
2.4 Hypotheses		49
2.4.1 Hypothesis 1: Perceived prosocial impact reduces performance gaming.		49

2.4.2 Hypothesis 2: Public managers who activate moral identity engage in less performance gaming compared to managers who do not.....	50
2.4.3 Hypotheses 3 & 4: 3) Perceived prosocial impact and gaming.....	51
2.4.4 Experimental research	52
2.5 Conclusion.....	53
III. ESSAY 2: EXPLAINING EFFORT SUBSTITUTION IN PERFORMANCE SYSTEMS: THE ROLE OF TASK DEMANDS AND MISSION ORIENTATION.....	54
3.1 Abstract	54
3.2 Introduction	55
3.3 Gaming of Performance Systems.....	57
3.4 Task Demands and Mission Orientation	63
3.5 Data and Methods.....	67
3.5.1 Research Context and Data.....	67
3.5.2 Measuring Effort Substitution	71
3.5.3 Independent Variables	74
3.6 Results	81
3.7 Discussion	84
3.8 Conclusion.....	89
IV. ESSAY 3: DOES PROSOCIAL IMPACT REDUCE PERFORMANCE DATA GAMING? THE ROLE OF DATA VISUALIZATIONS AND EXPERT-NOVICE DIFFERENCES	90
4.1 Abstract	90
4.2 Introduction.....	91
4.3 General Model of How Prosocial Impact Reduces Performance Gaming.....	93
4.4 Theoretical Adjustments for Expert-Novice Differences.....	98
4.5 Data and Methods.....	100
4.5.1 Capturing Performance Gaming.....	101
4.5.2 Treatments and Other Variables	103
4.5.3 Samples of Citizens and Professionals	104
4.6 Results	108
4.7 Discussion	112
4.8 Conclusion.....	115

V. CONCLUSION	118
5.1 Summary of Conclusion.....	118
5.2 Conclusion of Essay 1: Systematic Review	120
5.3 Conclusion of Essay 2: Effort Substitution.....	122
5.4 Conclusion of Essay 3: Prosocial Impact.....	123
5.5 Synthesized Conclusion	124
5.5.1 Other-Oriented Constructs: Toward a Theory of Performance Gaming	126
5.5.2 Contributions to theory and practice	130
REFERENCES	134
Appendix.....	158
VITA.....	180

LIST OF TABLES

TABLE	subheadings	PAGE
Table 1.	Factors of performance gaming and unethical behavior.....	29
Table 2.	Excluded Articles.....	37
Table 3:	Effort Substitution Measures and Factor Structure	73
Table 4:	Descriptive Statistics and Correlations.....	78
Table 5:	Acceleration gaming models: Unlinked performance measures	79
Table 6:	Graduation gaming models: Linked performance measures	80
Table 7:	Descriptive Statistics	107
Table 8:	The effect of prosocial impact treatments on performance gaming	109
Table 9:	Moderations of the effect of prosocial impact (school leaders)	111
Table 10.	Appendix 2.A: Treatment-demographics correlations.....	170
Table 11.	Appendix 2.B: The effect of prosocial impact (without controls).....	170
Table 12.	Appendix 2.C: Does Attention Matter?	171
Table 13.	Appendix 2.D: Student-Administrator Match Based on Race.....	172
Table 14:	Breusch-Pagan / Cook-Weisberg Test for Heteroscedasticity Grad Gaming.	174
Table 15:	Breusch-Pagan/Cook-Weisberg Test for Heteroskedasticity Acc. Gaming... ..	175
Table 16:	Graduation Gaming Model 1: Charter School.....	176
Table 17:	Graduation Gaming Model 2: Mission Specificity.....	177
Table 18:	Acceleration Gaming Model 1: Charter School	178
Table 19:	Acceleration Gaming Model 2: Mission Specificity	179

I. INTRODUCTION

1.1 Purpose of the Study

In this section, I will first explain why the topic of performance gaming is important to public administration. Then, I will explain the central concepts of performance gaming and the main variables of interest, which are the other-oriented constructs of mission orientation and prosocial impact. Finally, I will conclude this section by explaining the research question and discussing the purpose of the study.

1.1.1 Importance of the topic

When organizations measure their performance, and use these measures to make decisions, they reap many benefits (Hatry, 2007). In public organizations, the benefits include accountability to the public through performance reports, which provide transparency in how the government makes decisions. The performance literature offers a vast amount of empirical support for the benefits of performance systems (Moynihan & Beazley, 2016; Van de Walle, 2009; Van Dooren et al., 2015). But there are also many problems with performance systems, referred to as system dysfunctions. These include system design issues that fail to genuinely capture organizational performance (Heinrich & Marschke, 2010) and minimal or nonuse of performance information (Van Dooren & Van de Walle, 2008), where data collection is a mere box to be checked by managers with no consequences for decision-making. But most importantly, and the topic of this dissertation, is a dysfunction of performance systems known as performance gaming. This is a perverse form of information use (Moynihan, 2009), where the intent is to

distort performance measures, usually to meet a target. Performance gaming is important for both practical and theoretical reasons. From the practical perspective, the detrimental effects of performance gaming have been well-documented with storied examples, such as the hospital wait-times scandal that wreaked havoc in the U.S. Veteran's Affairs Department in 2014 or the cheating scandal in the public school systems of Atlanta and Washington, D.C., which occurred in 2009 and 2018, respectively. In the Veteran's Affairs Department scandal, public managers, under pressure to meet targets, manipulated hospital wait-times to make it appear that military veterans were receiving medical care in a timely manner (Oppel, Jr. & Shear, 2014). D.C. public school officials fudged graduation numbers to make it appear that high school students had earned diplomas, when in fact they had not (Jamison, 2018). In Atlanta Public Schools, administrators and teachers cheated on the state standardized test by erasing or changing answers (Blinder, 2015). In the New York City Police Department, whistleblowers complained about ticket quota systems that forced them to meet ticket performance targets (Rose, 2015). Many of the public servants in these cases faced criminal charges and some were convicted. The obvious consequences were a loss of public trust, an erosion of democratic accountability, and the most serious consequence: vulnerable groups not receiving the needed services. Thus, from a practical perspective, the issue of performance gaming is important because of its effects on the public and important government institutions.

For organizational theorists and management scholars, the issue of performance gaming is important because of implications for how organizations and managers make decisions. In most organizations, dysfunctional behaviors have generally been either

taboo or fallen under a gray area of ethical uncertainty, making it difficult for scholars to judiciously examine these behaviors or for practitioners to identify their actual effects on decision-making and organizational effectiveness. Studying performance gaming within managerial decision-making also contributes to our understanding of how organizations could enhance the benefits they provide to the public. To sum, the issue of performing gaming is important for public management theory, organizational behavior, and, in practice, has the potential to benefit the general public.

1.1.2 Definition of Central Concepts

The central concept that this dissertation examines is *performance gaming*, a behavior which involves generating positive performance data without achieving the actual objective behind the indicator (Kelman & Friedman, 2009; Moynihan, 2009; Smith, 1995). Performance gaming encompasses improving “measures in ways that are in conflict with the underlying or unmeasured goals of a program” (Moynihan, 2009, p. 593). In exploring the factors of performance gaming, this dissertation takes a different approach from previous research by looking at particular factors that are not found in performance system designs. These factors can be described as other-oriented constructs. These concepts are embedded in the idea of altruism, that is, helping others. Public organizations fit well into the conversation on other-oriented constructs because there is plenty of literature on public service motivation and the propensity for public administrators to have a stronger disposition for serving the public compared to their private counterparts (Pandey & Stazyk, 2008; Perry, 1996). Other-oriented constructs are centered on a concern for others (Bolino & Grant, 2016; Korsgaard et al., 1996).

Examples include prosociality, public service motivation, mission orientation, and organizational citizenship behavior. I explore two other-oriented constructs—prosociality and mission orientation—to determine whether they are likely to reduce performance gaming. *Prosociality* is a three-prong stream centered on how individuals help others through their work (Bolino & Grant, 2016). It is comprised of prosocial motives, prosocial behavior, and prosocial impact. This dissertation is concerned with prosocial impact, which is the experience individuals feel when they know someone has benefited from their work (Bolino & Grant, 2016). The second concept, mission orientation, serves as a motivation for organizational behavior and is based on purposes that are focused on serving others (Brewster & Cerdin, 2018). While all organizations have a mission, public and nonprofit organizations have missions that are generally focused on serving the public and providing for the welfare of others. I define *mission orientation* as the degree to which organizations are focused on their mission, a purpose that is generally prosocial (Knies & Leisink, 2018; Moore, 1995). These other-oriented constructs provide an alternative explanation for understanding the relationship between public managers in performance management systems and other stakeholders and how they use performance information. Rather than depending heavily on self-interest as the driving motivational force of performance information use, other-oriented constructs show that users of information are motivated by prosocial and mission-driven mechanisms.

1.1.3 Research question: What are the determinants of performance gaming?

The purpose of this dissertation is to identify predictors of performance gaming – why does this behavior occur and what can be done to minimize such behavior. This dissertation examines performance gaming in public organizations because public

organizations compared to private ones benefit and function differently (specifically along the lines of for-profit private enterprises). The difference also extends to the area of performance systems. Performance systems allow for the benefits of democratic accountability, transparency, and ultimately improved service for the public (Van Dooren et al., 2015). Understanding dysfunctions of performance system in the public sector has implications for historically disadvantaged groups, who become doubly disadvantaged under dysfunctional performance regimes (Soss et al., 2011).

1.2 State of the Literature and Research Gaps

In the performance management literature within public administration, performance gaming has generally been studied as a function of performance system flaws or system pressure (Courty & Marschke, 2007; Hood 2006). Public servants learn performance systems and then game the system based on knowledge of loopholes and ways to cut corners (Courty & Marschke, 2007). Often, there are targets to meet or ranking systems by which individuals or organizations are compared to one another, which creates pressure to perform or outperform other units (Bevan & Hood, 2006). These pressures and system loopholes often lead to performance gaming (Boyne & Chen, 2007). The public administration literature of performance gaming has not examined other-oriented constructs, such as prosociality and mission-orientation, and their role in reducing performance gaming. This dissertation fills this gap by drawing on other social science disciplines. Because performance gaming falls under the broad scope of

performance management, it is important to examine how this dissertation fits into this stream of literature.

1.2.1 Principal-Agent Theory and Performance Gaming

The major theory of performance management is principal-agent theory. Also called agency theory, it rests on the premise that agents (public managers, to use one context of government) and principals (elected officials) have differing perspectives and are motivated by self-interest (Moynihan, 2008). Seminal work for agency theory was first proposed by Barry Mitnick and Stephen Ross at around the same time in the early 1970s. Ross (1973) proposed the theory in terms of economics, describing the situation as an issue with incentives for an agent and how best to maximize those incentives. Therefore, Ross's theory rests on an optimal incentive system. Mitnick (1975) proposed an institutional perspective of agency theory, where institutions must manage and even adapt to the relationship of agents and principals. According to Mitnick, understanding of institutional structures are important to the principal-agent relationship. The origination of agency-theory on the behavior of actors within the firm itself rests with the influential work of Jensen & Meckling (1976), who proposed that agents do not maximize a firm's return, situation referred to as agency loss, and incentives must be included by the principal in order to alleviate this loss.

As the theory developed, information became critical to determining incentive systems. Agents will use information asymmetry to keep principals in the dark, being selective in the provision of information. Therefore, principals must invest in monitoring systems and to contract on the agent's behavior or the outcome that the agent produces (Eisenhardt, 1989). A key assumption of information asymmetry is that agents are self-

interested, thus the need for monitoring systems and incentives (Perrow, 1986). This is a very interesting point of the theory for organizational scholars and particularly those who study public organizations because of the nature of serving the public.

Agency theory has generally neglected the fact that altruistic values may alter the relationship between principals and agents. However, there is increasing research to suggest that other-oriented constructs, such as prosociality, the concept of an individual benefiting others, usually through the individual's work, may play an important role in the principal-agent relationship (Gailmard & Patty, 2007; Heinrich & Marschke, 2010). My theory expands on the principal-agent framework, by saying that agents are not entirely utility maximizers. Some agents, particularly in mission-driven organizations (or from their own predisposition) are prosocial maximizers. To some extent, they are looking to maximize their own utility, but to some extent they are looking to help others. So, this is an enhanced version of the principal-agent theory. Critics of the traditional form of agency theory have proposed that when principals depend on incentivized performance management systems, they are unable to prevent gaming the incentive measures and they also do not account for other motivations not captured by the incentive systems (Le Grand, 2006; Moynihan, 2010; Weibel, Rost, & Osterloth, 2010).

My dissertation tests this expansive approach to principal-agent theory and addresses the criticisms of the traditional theory. Empirical tests of the alternative agency theories, which incorporate altruism instead of self-interest, include bureaucrats who are motivated by policy (Gailmard & Patty, 2007; Moynihan et al., 2012), which examines prosociality as a variable that affects performance information use. In the next section, I address mission orientation as an other-oriented construct.

1.2.2 The Theory of Mission Orientation

Along with prosociality, mission-orientation is part of the expansive view to agency theory explained in this dissertation. The literature on mission-orientation has quite meaningful empirical findings and theory building for organizational behavior and performance. An organization's mission statement offers a window into understanding the purpose and direction of an organization as seen by its founders and leadership (Weiss & Piderit, 1999). Often using content analysis of mission statements and vision statements, researchers have concluded that mission statements can shape organizational behavior (Weiss & Piderit, 1999; Garrow & Gursky, 2013), though other studies have shown a disconnect between mission and vision statements and day-to-day functions (Abelman et al., 2007; Schafft & Biddle, 2013). Researchers have also made correlations between mission and performance outputs (Boyne & Chen, 2007; Chun & Rainey, 2005; Garrow & Gursky, 2013), and have also made distinctions between different types of mission-orientation, finding a philosophical distinction to be an important type of mission-orientation (Boerema, 2006; Henig et al., 2005).

Probably the most significant piece of knowledge within the literature on mission-orientation is how scholars have determined the types of mission. Mission-oriented organizations can be distinguished based on their distinct philosophical beliefs. Albert Boerema (2006) called this a *school's distinctive belief* and he found that among private schools, there was considerable diversity along this dimension compared to three other themes that were studied. Among charter schools, Brown et al. (2004) identified one of two strands that fit under the scope of a specific mission-orientation. They described autonomy and decentralized control as the first strand of mission-oriented schools. And

they identified a specific philosophical approach as the second strand. In the distinct philosophical mission-orientation strand, teachers and families are united by a mission based on a “set of animating values” that may be specific to the process of delivering education, such as a “back-to-the-basics approach—but they may also have their roots in racial, ethnic, or religious identity” (Brown et al., 2004, p. 1038). The typology of a distinct philosophical mission-orientation is empirically tested by Henig et al. (2005), and they conclude with two broad types of organizations: market-oriented and mission-oriented. They find limited differences among the various types of mission-oriented organizations, but a distinct difference between market and what they call more mission or purpose-driven organizations.

From their research, a mission-orientation (or a market one for that matter) is based on the perspectives of the founders – the philosophy, values, and traditions of the founders are formulated in the mission. In this dissertation, I expand on the founder’s philosophy as a mission-orientation type and also look at who the specific clients are in identifying a specific mission from a general mission. This is because who or what is being served are important dimensions of a mission (Brown et al., 2004; Garrow & Gursky, 2013). Henig et al. (2005) identified differences in mission-oriented organizations and found that service to specific high-need populations were a defining feature of community-based mission-oriented organizations. So, even in their study, understanding the specific target groups are important dimensions of mission-orientation.

1.2.3 Conclusion of Theoretical Foundation

Prosociality and mission-orientation are other-oriented constructs that provide alternative explanations to principal-agent theory, the traditional theory that explains performance management and information use. For public organizations, the theories of prosociality and mission-orientation are consistent with other similar theories, such as public service motivation (Perry, Hondeghem, & Wise 2010). In public organizations, individuals are said to have greater dispositions for motivations to serve the public (Perry & Wise 1990). This section provided a detailed explanation of the theoretical underpinning of my dissertation. In the following sections, I summarize my three essays and how I tested these theories as explanation for dysfunctional uses of performance information.

1.3 Argument of the Dissertation

The main argument of my dissertation is that other-oriented constructs, such as mission orientation and prosocial impact, are inversely associated with performance gaming, meaning that they tend to decrease the undesirable effects of the factors that lead to this dysfunction. This dissertation points scholars of public management and organizational behavior into a different perspective on dysfunctional use of performance information. The models that we use to understand performance systems should include more than just system factors and organizational variables, but also these other-oriented

constructs. This section explains these other-oriented constructs in detail and the foundation for the main argument of my dissertation.

1.3.1 The Behavioral Perspective for Understanding Performance Gaming

Traditionally, public administration scholars have focused on system factors as antecedents of performance gaming. Targets and ranking systems are influential determinants of gaming behavior (Bevan & Hood, 2006), as well as system loopholes that allow performance systems to be learned and gamed (Courty & Marschke, 2007). Other predictors include the maturity level of the performance system (Courty et al., 2005), systemic issues or flaws in system measurements (Heckman, Heinrich, & Smith, 2002; Kelman & Friedman, 2009), and organizational variables such as task demands (Bothe & Meier, 2000). In other social science disciplines, predictors of performance gaming and unethical behavior have been at the individual level and have taken a more behavioral perspective. These other social science disciplines have shed light on antecedents that the public administration literature should give more attention to.

Other-oriented constructs, such as mission-orientation and prosociality, are avenues that may provide a vast array of knowledge in understanding organizational behavior that the public administration has overlooked. This dissertation argues that the performance gaming literature is void of these antecedents and can benefit from this behavioral perspective and the micro-level processes that lead to decision-making in performance systems.

1.3.2 Mission-Orientation in Public Organizations

All organizations have missions or some purpose for which they exist (Brewster & Cerdin, 2018), but public and nonprofit organizations will differ from private, for-profit firms in the premise of their mission. On the one hand, for-profit organizations will generally have a profit motive for their mission, while nonprofit and public organizations will have a purpose of serving some specific goal or the public interest. The term, *mission*, in this dissertation is based on the definition of mission pertaining to public and nonprofit organizations. Mission in this sense is a purpose that is other-oriented, or serving someone other than oneself. The mission-orientation examined in this dissertation is done at the organizational level and is a measure of how inclined organizations are to the idea of serving others. The premise for this argument is that research has shown strong prosocial motives give individuals and teams a strong sense of purpose and allow them to lean to behave prosocially, behavior done for the benefit of others (Korsgaard et al., 1997).

1.3.3 Prosocial Impact and Visualizations of Performance Data

The second other-oriented construct that this dissertation examines is prosocial impact, which is part of a three-prong stream that centers on helping others. The three-prongs are prosocial motives, prosocial behavior, and prosocial impact. Prosocial impact is the experience that individuals receive from knowing their work has benefited someone else (Bolino & Grant, 2016). I argue that prosocial impact and individual characteristics (such as experience and attention to information) frame how individuals visualize and process impact information, which in turn reduce performance gaming. The basis for this

argument is that previous studies have examined the framing effects of performance data and prosocial impact (Belle, 2013a; Grant 2008a; Olsen, 2017).

Studies have shown that prosociality has many positive outcomes, such as increased productivity among workers and higher levels of motivation, as well as greater accuracy in work. Prosocial behavior is tied to the idea of working for the public interest (Wright et al., 2016) and if performance gaming prevents individuals from, high levels of prosocial impact will likely lead to reductions in dysfunction behavior. This theory fills a gap in the performance management literature overall because of the need to better explain principal-agent theory. Researchers have called for an expanded view of agency theory that explains how individuals use information in performance systems (Kettl, 1997; Heinrich & Marschke, 2010). Prosociality is an other-oriented construct that can help explain such behavior such as performance gaming. And it also emphasizes the importance of individual actors in the agency model, because individual characteristics are still vital to explaining gaming behavior.

The complete argument is that prosocial impact, as an interaction with other characteristics (attention to data and experience with data), explains performance gaming. So, prosociality makes up the second of the two other-oriented constructs that predict performance gaming. The next section sums of the main argument for this dissertation, situating the argument in the context of the behavioral perspective.

1.3.4 Summary of the Main Argument

This dissertation examines performance gaming and takes an alternative approach to studying factors that can mitigate performance gaming. In particular, I take a behavioral perspective by studying two types of other-oriented constructs: mission-

orientation and prosocial impact and how these can reduce performance gaming in public organizations. The basis of my argument comes from a literature review of the performance gaming and unethical behavior literature of other social science disciplines and public administration. In my comparative review, I learn that public administration has largely focused on technical and system factors as determinants of performance gaming. Other social science disciplines have mainly focused on individual characteristics. So, I argue in my literature review that there is an opportunity for new directions in studying performance gaming where we can take a behavioral approach and look at other-oriented constructs. This fits in with research that has argued for an altruistic explanation of the principal-agent theory (Gailmard & Patty, 2007; Moynihan et al., 2012), where agents are not motivated by self-interest, but a desire to serve others. This is not entirely new to public administration, in the sense that scholars have long observed the benefits of these constructs, such as organizational citizenship behavior (Vigoda-Gadot & Beeri, 2011) and public service motivation (Perry & Wise, 1990). But it is novel in the study of performance systems and is part of a new wave of public administration scholarship that is taking on a behavioral perspective (Marvel, 2015; Nielsen & Moynihan, 2017; Olsen, 2015). This behavioral approach of public administration looks at the micro-level processes, as well as cognitive constructs “and it draws strongly both from psychological theories and from the experimental approach favored by that discipline” (Moynihan, 2018, p.1). This dissertation is part of this wave of behavioral perspective and makes the argument that performance gaming can be mitigated by such factors.

1.4 Approach of the Dissertation

This dissertation consists of three essays: 1) a literature review 2) a longitudinal, observational study and 3) an experimental study (the observational paper and literature review have been published in peer review journals). I chose this three-essay format because it allowed me to conduct multiple studies that were part of the overall research question and thus made for a stronger overall argument. The three essays have not only been reviewed by my dissertation committee, but also gone through the blind review process of academic journals, thus allowing additional experts in the field to review. This is related to the triangulation method of research that allows for increased reliability when the same phenomenon is studied from several angles and involves review by multiple researchers. “Through triangulation (using data sources, methods, theories, or researchers), the validity of specific knowledge claims is argued to be more robust” (Marshall & Rossman, 2016, p. 48).

Each chapter of essays is different in its sample, unit of analysis, method, and specific research question studied. But each essay is similar in that it is part of the broader and main argument of the dissertation. For example, the unit of analysis for the first essay (the literature review) is journal articles and the method is a systematic review. The second essay focuses on public organizations (schools) as the unit of analysis and uses pooled OLS regression as the method. The third essay has public managers and a general population of individuals as the unit of analysis. The method is a randomized experimental approach, using OLS regression. The overlap and similarities are that each study focuses on performance gaming as the phenomenon of interest, though each essay looks at different types of performance gaming. For example, the first essay looks at

various forms of performance gaming and unethical behavior, across different social science disciplines. The second and third essay looks at specific types of performance gaming, effort substitution and cream skimming, respectively. The main overlap of all the articles is that it looks at behavioral and other-oriented characteristics as the main independent variables of interest. This makes up the chief argument of the dissertation – that other-oriented constructs mitigate performance gaming.

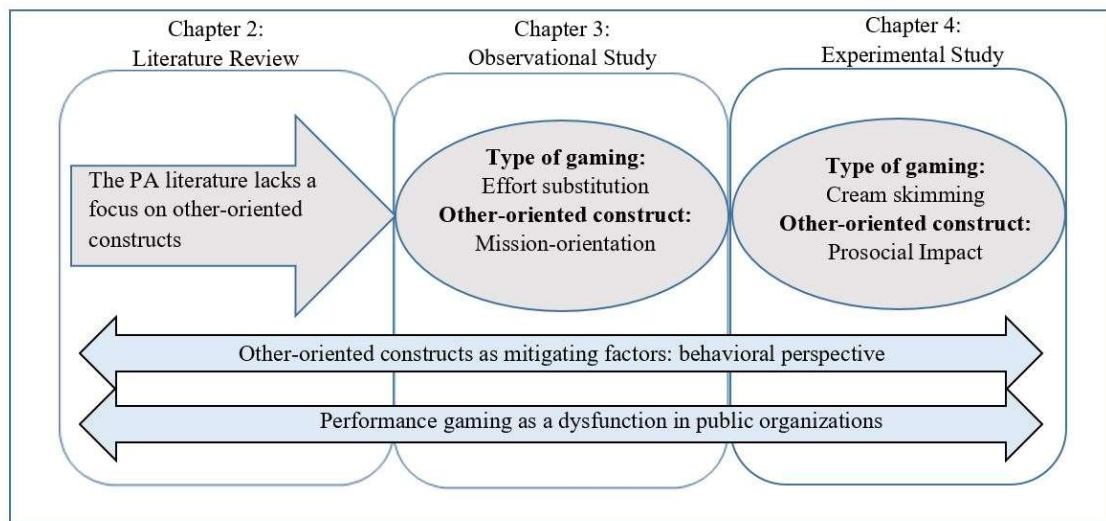
I used several methodological approaches for each of the three essays. The first essay uses the methodological approach of a systematic review, with two criteria for articles to be included in the review: 1) an empirical study and 2) performance gaming or cheating behavior was the outcome variable. This approach was necessary because I wanted to learn what antecedents had been empirically studied for performance gaming. The second essay used publicly available data on 64 high schools in one specific school district over a five-year period (2012-2016). This approach was chosen because schools have very mature performance systems and focusing on one school district allowed for separating system effects from organizational-level variables. The years were chosen because they were the most recent years for which data were available at the time of the study. The methodology for the third study involved a randomized, experimental study consisting of a general population sample and a sample of public managers. An experimental study was a necessary approach because I wanted to isolate the treatment (impact information) and see whether this treatment had an effect on how individuals gamed performance.

Public administrators in the education sector have a long-standing experience with performance systems. In Florida, for example, school districts have used performance

systems since the 1990s, even before the state created its school accountability system (Florida Department of Education, 2017). Therefore, public administrators are familiar with the performance pressures presented in the experimental study. Schools in the district I examined were ideal for the observational study because the school district was a pioneer of performance monitoring in the state of Florida, as evidenced by performance data available on their web pages, which predates the state's performance system.

Figure 1 is a visual representation of my dissertation approach and shows how each essay chapter is linked and how each chapter supports the overall research question and the dissertation's main argument. The literature review chapter provides the foundational support for the subsequent chapters and the dissertation's main argument. This main argument is that public administration should expand its focus and look at other-oriented constructs as mitigating performance gaming. Across all three chapters, two key themes appear 1) performance gaming is the key variable of interest and 2) other-oriented constructs are mitigating factors that provide for a behavioral perspective of examining performance system dysfunctions. Each chapter then decomposes these themes, with the two major studies of my dissertation focusing on a specific type of gaming and a specific other-oriented construct. In the next section, I will summarize each chapter of my dissertation. Then, I will conclude the introduction of the dissertation and introduce each essay and the essay themselves. At the end of the essays, I will provide a conclusion and bring each of the essays together for a summative conclusion on how each chapter is linked.

Figure 1. The relationship between chapters



1.5 Explanation of the Three Essays

My dissertation comprises of three essays, which include a systematic literature review of performance gaming and two empirical studies. The systematic review compares the performance gaming literature of public administration to that of other social science disciplines in order to draw lessons for research. The second empirical essay is a randomized, experimental research that consists of two studies. Study 1 uses a general population sample and study 2 uses a sample of public managers, where I study

the effects of prosocial impact on performance gaming through different visualizations of performance data. The final section of my dissertation is the conclusion where I conduct a synthesis of the two empirical essays to build the case for the importance of other-oriented constructs. Using the findings from my empirical studies, I provide a response to the problem statement described earlier: public organizations are plagued by dysfunctional consequences of performance systems, such as performance gaming and it is important to understand what factors can minimize this behavior. In the three following sections, I will provide a more detailed summary of my three essays.

The primary problem that my dissertation addresses is understanding the factors that mitigate dysfunctional consequence of performance systems, particularly performance gaming. I observe this behavior in public organizations as oppose to private or quasi-public entities. The publicness demarcation is an important one because of two main reasons: 1) public organizations benefit from appropriately functioning performance systems in a very unique way compared to private firms (Moynihan, 2008; Van Dooren, & Van de Walle, 2008), and 2) I take a very different set of performance gaming antecedents compared to previous research on public organizations, where the antecedents have been system-related drivers (Bevan & Hood, 2006; Bohte & Meier, 2000; Heinrich & Marschke, 2010).

First, public organizations are unique in the benefits they receive from performance systems primarily because public agencies are induced to “become more efficient, responsive, and accountable to the public” (Heckman et al., 2011, p.1). This benefit of democratic accountability comes through transparency in performance reports

and responsiveness in budget allocations (Heckman et al., 2011). This ultimately results in the appropriate allocation of resources and informative decision-making for the public's benefit. These benefits are based on the assumption that performance measurement in such systems are accurate and aligned with organizational goals. When this is compromised because of dysfunctional behavior, then the benefits of democratic accountability and the public welfare are compromised (Yang, 2009). The second reason why my dissertation is unique for public organizations is because it is different from previous research in public administration.

In most of the performance gaming literature in public administration, the antecedent of performance gaming are system-related factors (Bevan & Hood, 2006; Radin, 2000). These include targets, where the pressure to reach a preset threshold makes gaming the system attractive (Bevan & Hood, 2006). System-related factors also include system dynamism, where static systems can easily be learned and then gamed (Heinrich, 2007). Incentivized systems have been linked to performance gaming because performance outcomes tied to rewards result in pressures to attain these outcomes (Heinrich & Marschke, 2010; Kelman & Friedman, 2009). These are the antecedents that scholars have usually observed when they have examined performance gaming in public organizations. However, in other social science fields, the focus has been a more behavioral perspective that settles on individual drivers as the antecedents of performance gaming (Gino et al., 2009; Schweitzer & Hsee, 2002). The unit of analysis in these other disciplines is often individuals and the methodical approach is usually experimental. Public organizations are not usually examined in these other fields. Usually graduate students are used or participants from the general public.

My dissertation draws lessons from these other fields and focuses on an individual and behavioral perspective with public managers and public organizations as the units of analysis. I examine other-oriented constructs, such as prosociality and mission orientation as antecedents. I also examine interactive effects of these variables with organizational variables (such as task demands) and individual traits (attention to performance data and experience levels). In my dissertation, I contribute to theory-building in organizational behavior by focusing on the human side to how organizations and individuals behave and in particular how they use information in dysfunctional ways. My dissertation lays the foundation for exploring these other-oriented constructs and how they influence performance gaming.

II. ESSAY 1: PERFORMANCE GAMING: A SYSTEMATIC REVIEW OF THE
LITERATURE IN PUBLIC ADMINISTRATION

Emerald Publishing Limited owns the copyright for Essay 1 and requests that the following statement appear on the first page of the reprinted article:

This article is © Emerald Publishing Limited and permission has been granted for this version to appear here <https://doi.org/10.1108/IJPSM-07-2019-0191>. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Publishing Limited.

2.1 Introduction

Traditionally, the performance management literature has focused on ways to improve the performance system in the technical sense. As Radnor (2008a) puts it, “a systems view ... promotes appropriate behavior, provides a mechanism for accountability and control, and creates a mechanism for intervention and learning” (p.95). But this performance management doctrine made faulty assumptions that did not account for what really happens in performance systems (Moynihan, 2008). One of the assumptions was that performance measurement systems will accurately lead to organizational goals and effective organizations. This assumption does not account for natural system consequences that undermine underlying organizational goals. Over the years, the performance management literature has given this dysfunctional behavior (performance gaming) considerable attention, mainly explaining why performance dysfunctions occur from a system perspective (Courty & Marschke, 2007; Hood, 2006; Smith, 1995; van Thiel & Leeuw, 2002). This paper picks up at this point and expands on factors that contribute to dysfunctions of performance systems. While public administration has focused on system factors that influence performance gaming, this paper suggests that public administration should take a behavioral approach, focusing on individual-level variables, such as moral identity and other-oriented constructs. This conclusion comes from a comparative analysis of other social science fields and the unethical behavior literature. To summarize in a less technical manner, this paper suggests a path forward for understanding why public organizations and individuals within those organizations engage in performance gaming. It uses research from other social science disciplines,

which suggests that we can research performance gaming by examining behavioral traits, such as being prosocial and examining the purpose of who and why organizations serve.

2.2 Methods

The systematic review compares studies in public administration to those of other social science fields. Journal articles were limited to empirical studies, using Google Scholar, Web of Science, and ProQuest Central. The systematic review covers years ranging from 1990 through 2019. This span of years was chosen because it coincides with the rise of New Public Management and the era of research into performance systems. During the 1990s, the United States federal government implemented the Government Performance and Results Act (GPRA) and many would argue that this ushered in the era of governance by performance management (Moynihan, 2008). In this review, a total of 81 articles were found and 51 empirical studies were deemed relevant (based on criteria explained in the next section). These articles were from academic journals in both the public administration field, as well other social science disciplines.

There were 26 articles that focused on individual factors (one from the public administration literature, 6 from economics, 8 from business, and 11 from psychology). In addition, 14 articles focused on organizational factors (3 of these were from the public administration literature, 3 from economics, 5 from business, 2 from psychology, and one from education). For system factors, there were a total of 20 articles (6 of these were public administration papers, 6 from economics, 3 from business, 4 from psychology, and one from education). These categories were not exclusive as some articles focused on drivers from more than one category. The number of articles from the public

administration and the other social science disciplines are listed when each category of drivers is discussed below.

2.2.1 Systematic literature review

An article was deemed relevant if it was 1) an empirical study and 2) performance gaming or cheating behavior was the outcome variable. Out of the 81 articles surmised from the search, 51 were deemed relevant by meeting these criteria. To obtain the articles, the systematic review featured two major avenues. The first avenue was a database keyword search, using the aforementioned databases. Key words, such as “performance gaming”, “dysfunctional consequences” and “performance information” were used. Also, terms such as “dishonesty”, “unethical behavior”, and “cheating” were also used. The second avenue of the search process was to employ a snowball technique, used by Belle & Cantarelli (2017a) in their meta-analytic review. In the snowball technique, a researcher looks at foundational studies in the literature and then searches for references and articles that have cited the papers, being sure that each reference is theoretically linked to the foundational article. In this study, foundational studies in the performance gaming literature for public administration were examined (Bohte & Meier, 2000; Smith, 1995; Hood, 2006) and unethical and performance gaming research that are foundational in other disciplines were examined (Barkan et al., 2012; Becker, 1968; Hill & Kochendorfer, 1969). If the articles were linked to performance gaming and based on the relevancy criteria, they were included. If they did not meet these criteria, they were excluded.

Most of the drivers fell under three major categories: performance systems factors, individual factors, and organizational factors. The papers were then divided among these factors. To code articles, each antecedent factor was defined. For the system category, articles would need to have antecedents associated with the performance system itself, such as the type of system or the maturity level of the performance system. For articles to be included in the organizational category, organizational variables, such as resources or political control would need to be labeled as antecedents. And finally, for the individual factor categories, articles were included if the antecedents were behavioral actions or psychological traits of the individual. Because some articles studied more than one factor, some articles appear in more than one category. Articles that appeared in the public administration journals were coded under the public administration category and articles from journals outside public administration were coded as “NO” for not being from the discipline. The PRISMA flow diagram below (Figure 2) illustrates how articles were systematically included in the review. PRISMA flow diagrams are used by scholars to improve reporting of systematic reviews (Liberati et al., 2009). In the PRISMA diagram, 8 articles overlap into more than one category. So, although the total number of relevant articles is 51, the sum of the total articles from the categories is 59. Table 1 shows all the articles that were included in the systematic literature review, including the factors that lead to performance gaming, the factor category, and whether the article is from public administration. Table 2 includes all the excluded articles, including the reasons for exclusion.

Figure 2 PRISMA Flow Diagram

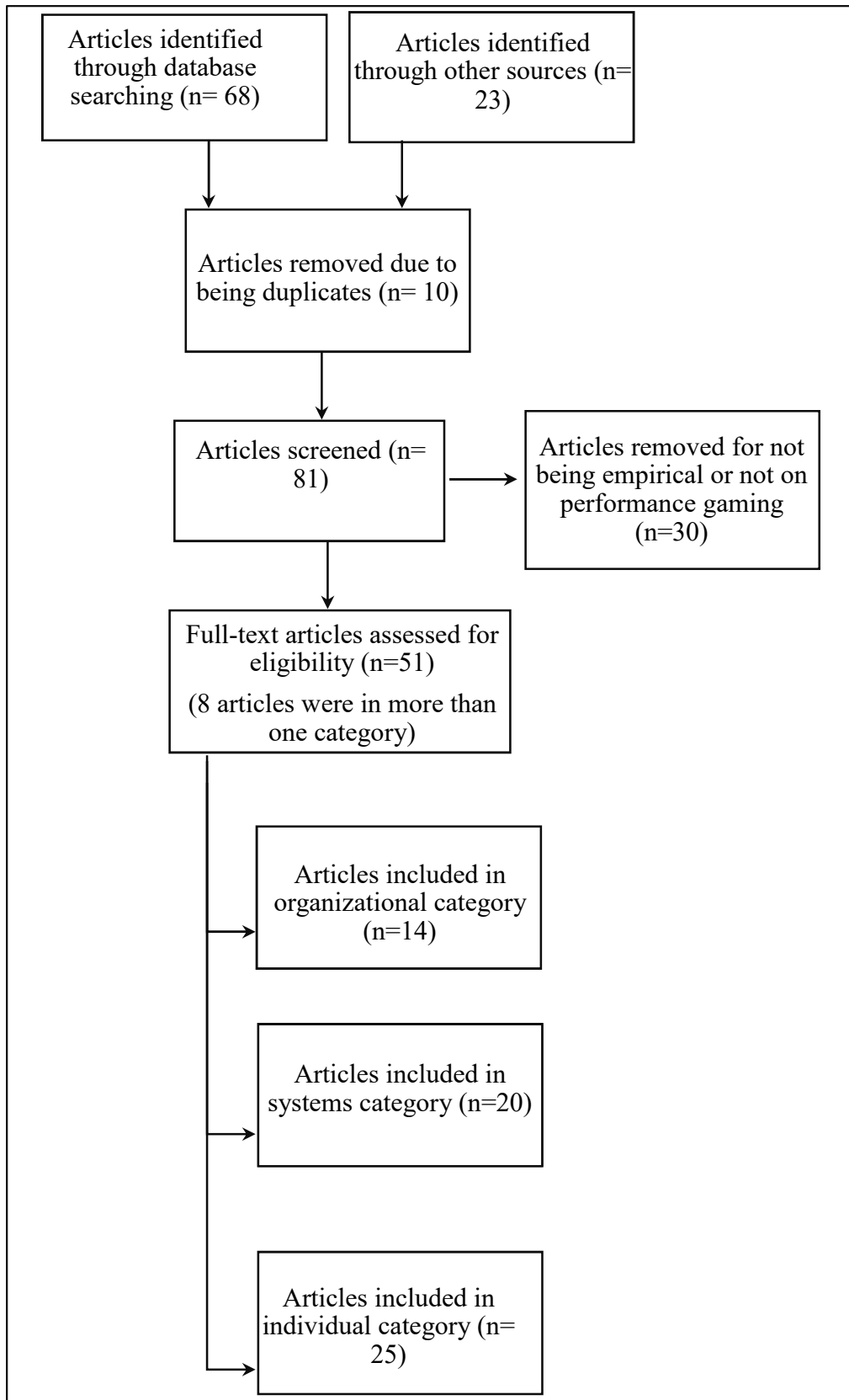


Table 1. Factors of performance gaming and unethical behavior

Authors	Title of article	Journal	Factors that lead to gaming	Factor Category: System, Organization, or Individual	Public Administration
Abeler et al., 2014	Representative Evidence on Lying Costs.	Journal of Public Economics	Self-image, Self-Engagement, or Self-Consciousness or Moral Identity,	Individual	NO
Andrade and Ho, 2009	Gaming Emotions in Social Interactions.	Journal of Consumer Research	Self-Interest, incentives	Individual	NO
Aquino et al., 2009	Testing a Social-Cognitive Model of Moral Behavior: The Interactive Influence of Situations and Moral Identity Centrality	Journal of Personality and Social Psychology	Self-image, Self-Engagement, or Self-Consciousness or Moral Identity, Situational Factor x Moral Identity	Individual	NO
Ashforth and Anand, 2003	The Normalization of Corruption in Organizations	Research in Organizational Behavior	Self-Justification, Organizational Culture	Individual, Organization	NO
Barkan et al., 2012	The Pot Calling the Kettle Black: Distancing Response to Ethical Dissonance	Journal of Experimental Psychology	Self-image, Self-Engagement, or Self-Consciousness or Moral Identity	Individual	NO

Belot and Schroder, 2013	Sloppy work, lies and theft: A novel experimental design to study counterproductive behavior	Journal of Economic Behavior and Organization	Incentives, competition	System	NO
Bevan and Hood, 2006	What's Measured is What Matters: Targets and Gaming in the English Public Health Care System	Public Administration	Targets	System	PA
Bloodgood et al., 2010	Ethics Instruction and the Perceived Acceptability of Cheating.	Journal of Business Ethics	Self-Interest	Individual	NO
Bohte and Meier, 2000	Goal Displacement: Assessing the Motivation for Organizational Cheating.	Public Administration Review	Task Demands, Resources	Organization	PA
Burton and Near, 1995	Estimating the Incidence of Wrongdoing and Whistle Blowing: Results of a Study Using Randomized Response Technique.	Journal of Business Ethics	Visibility/Monitoring	Individual	NO
Cadsby et al. 2010	In-Group Favoritism and Moral Decision-Making	Journal of Economic Behavior & Organization	Incentives, Organizational Culture	Organization	NO

Carhart et al., 2002	Leaning for the Tape: Evidence of Gaming Behavior in Equity Mutual Funds	The Journal of Finance	Incentives	System	NO
Cojoc and Stoian, 2014	Dishonesty and charitable behavior	Experimental Economics	Temporal Discounting	Individual	NO
Courty and Marschke, 2007	Making Government Accountable: Lessons from a Federal Job Training Program	Public Administration Review	Incentives, Feedback loop	System	PA
Erat and Gneezy, 2012	White Lies	Management Science	Prosocial motives/altruism	Individual	NO
Fischbacher and Föllmi-Heusi, 2013	Lies in Disguise-An Experimental Study on Cheating.	Journal of the European Economic Association	Self-image, Self-Engagement, or Self-Consciousness or Moral Identity	Individual	NO
Gino et al., 2009	Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel	Psychological Science	Organizational Culture	Organization	NO
Gino et al., 2013	Self-Serving Altruism? The Lure of Unethical Actions that Benefit Others	Journal of Economic Behavior & Organization	Incentives, self-interest, Prosocial motives/altruism,	System, Individual	NO
Glac et al., 2012	Conflict in Roles: Lying to the In-Group Versus the Out-Group in Negotiations	Business & Society	Visibility/Monitoring	Organization	NO
Grimes and Rezek, 2005	The Determinants of Cheating by High School Economics	International Review of	Self-Justification, Organizational Culture	Individual, Organization	NO

	Students: A Comparative Study of Academic Dishonesty in the Transitional Economies	Economics Education			
Grolleau et al, 2016	Cheating and Loss Aversion: Do People Cheat More to Avoid a Loss?	Management Science	Lie-specific mechanisms	Individual	NO
Hilbig and Hessler, 2013	What Lies Beneath: How the Distance Between Truth and Lie Drives Dishonesty	Journal of Experimental Social Psychology	Lie-specific mechanisms	Individual	NO
Hood, 2006	Gaming in Targetworld: The Targets Approach to Managing British Public Services.	Public Administration Review	Targets, Incentives	System	PA
Hurkens and Kartik, 2009	Would I lie to you? On social preferences and lying aversion.	Experimental Economics	Incentives	System	NO
Juenke, 2005	Management Tenure and Network Time: How Experience Affects Bureaucratic Dynamics	Journal of Public Administration Research and Theory	experience	Individual	PA
Kaufmann et al., 2005	Ethical Distancing: Rationalizing Violations of Organizational Norms	Business & Professional Ethics Journal	Self-Justification	Individual	NO

Kelman and 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	target, system, feedback loop,	System	PA
Larkin, 2014;	The Cost of High-Powered Incentives: Employee Gaming in Enterprise Software Sales	Journal of Labor Economics	Incentives	System	NO
Mazar et al., 2008	The Dishonesty of Honest People: A Theory Of Self-Concept Maintenance	Journal of Marketing Research	Incentives, Organizational Culture, Self-image, Self-Engagement, or Self-Consciousness or Moral Identity	System, Organization, Individual	NO
McCabe and Trevino, 1993	Academic Dishonesty: Honor Codes and Other Contextual Influences	Journal of Higher Education	Organizational Culture, incentives	Organization, System	NO
Mead et al., 2009	Too Tired to Tell The Truth: Self-control Resource Depletion And Dishonesty	Journal of Experimental Social Psychology	Cognitive depletion	Individual	NO

Mitchell et al., 2018	Cheating under pressure: A self-protection model of workplace cheating behavior.	Journal of Applied Psychology	Targets, anger, Self-Interest	System, Individual	NO
Nagin et al., 2002	Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment.	The American Economic Review	Visibility/Monitoring	Organization	NO
Ottaviani and Squintani, 2006	Naive Audience and Communication Bias	International Journal of Game Theory	Targets	System	NO
Radnor, 2008b	Muddled, Massaging, Manoeuvring Or Manipulated? A Typology of Organizational Gaming	International Journal of Productivity and Performance Management	Targets, task demands	System, Organization	PA
Rigdon and D'Esterre, 2015	The Effects of Competition on the Nature of Cheating Behavior	Southern Economic Journal	Competition	System	NO
Schindler and Pfattheicher, 2017	The Frame of the Game: Loss-Framing Increases Dishonest Behavior.	Journal of Experimental Social Psychology	Lie-specific mechanisms	Individual	NO

Schocket & Burghardt, 2008	Do Job Corps Performance Measures Track Program Impacts?	Journal of Policy Analysis and Management	Targets,	System	PA
Schweitzer and Hsee, 2002	Stretching the Truth: Elastic Justification and Motivated Communication of Uncertain Information	Journal of Risk and Uncertainty	Self-Justification, competition	Individual	NO
Schwieren and Weichselbaumer, 2010	Does Competition Enhance Performance or Cheating? A Laboratory Experiment.	Journal of Economic Psychology	Competition	System	NO
Shalvi, 2012	Dishonestly increasing the likelihood of winning	Judgment and Decision Making	Situational Factor x Moral Identity	Individual	NO
Shalvi et al., 2011	Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior	Organizational Behavior and Human Decision Processes	Self-Justification	Individual	NO
Tenbrunsel, 1998	Misrepresentation and Expectation of Misrepresentation in an Ethical Dilemma: The Role of Incentives and Temptation	Academy of Management Journal	Incentives, organizational culture	System, Organization	NO
Utikal and Fischbacher, 2013	Disadvantageous Lies in Individual Decisions	Journal of Economic	Self-image, Self-Engagement, or Self-	Individual	NO

		Behavior & Organization	Consciousness or Moral Identity		
Vanberg, 2008	Why do people keep their promise? An experimental test of two explanations	Econometrica	Self-image, Self-Engagement, or Self-Consciousness or Moral Identity	Individual	NO
Vincent et al., 2013	Stretching the Moral Gray Zone: Positive Affect, Moral Disengagement, and Dishonesty	Psychological Science	Temporal Discounting, moral image	Individual	NO
Welsh and Ordonez, 2014	Conscience without Cognition: The Effects of Subconscious Priming on Ethical Behavior.	Academy of Management Journal	Visibility/Monitoring, moral image	Organization	NO
Wiltermuth, 2011	Cheating More When the Spoils Are Split.	Organizational Behavior and Human Decision Processes	Incentives, Prosocial motives/altruism	System	NO
Yang, 2009	Examining Perceived Honest Performance Reporting by Public Organizations: Bureaucratic Politics and Organizational Practice	Journal of Public Administration Research and Theory	organizational culture	Organization	PA
Zhang et al., 2008	Earnings Manipulation and	The Academy of	Incentives	System	NO

	Stock-Based Incentive Misalignment.	Management Journal			
Zhong et al., 2010	Good Lamps Are the Best Police: Darkness Increases Dishonesty and Self-interested Behavior	Psychological Science	Visibility/Monitoring	Organization	NO

Table 2. Excluded Articles

Authors	Title of article	Journal	Reason for exclusion	Public Administration
Alge et al., 2006	An Identity-Based Model of Organizational Monitoring: Integrating Information Privacy and Organizational Justice	Research in Personnel and Human Resources Management	theory paper	NO
Ayal et al., 2015	Three Principles to Revise People's Unethical Behavior	Perspectives on Psychological Science	theory paper	NO
Benabou and Tirole, 2011	Identity, morals, and taboos: Beliefs as assets	Quarterly Journal of Economics	theory paper	NO
Bevan and Hamblin, 2008	Hitting and missing targets by ambulance services for emergency calls: effects of different systems of performance measurement within the UK		not an empirical paper	
Burgess and Ratto, 2003	The Role of Incentives in the Public Sector: Issues and Evidence	Oxford Review of Economic Policy	theory paper	PA

Crossan et al., 2013	In search of virtue: The role of virtues, values and character strengths in ethical decision making	Journal of Business Ethics	theory paper	NO
Cuganesan et al., 2014	The Riskiness of Public Sector Performance Measurement: A Review and Research Agenda	Financial Accountability and Management	theory paper	PA
Frey et al., 2013	Organizational Control Systems and Pay-for-Performance in the Public Service	Organization Studies	dv is negative outcomes and not necessarily gaming in particular	PA
Gibbs et al., 2009	Performance Measure Properties and Incentive System Design	Industrial Relations: A Journal of Economy and Society	dv is incentive systems and not necessarily gaming in particular	NO
Hannah et al., 2011	Moral maturation and moral conation: A capacity approach to explaining moral thought and action	Academy of Management Review	theory paper	NO
Hood, 2012	Public management by numbers as a performance-enhancing drug: two Hypotheses	Public Administration Review	not an obs or experim paper, but topic is gaming	PA
Hood and Dixon, 2010	The Political Payoff from Performance Target Systems: No-Brainer or No-Gainer?	Journal of Public Administration Research and Theory	dv is negative outcomes and not necessarily gaming in particular	PA
Hruschka et al., 2014	Impartial institutions, pathogen stress and the expanding social network	Human Nature	redundant paper - studies cheating on behalf of in-group members	NO

				(Cadsby et al, 2015)	
Hyman, 2001	Health Care Fraud and Abuse: Market Change, Social Norms, and the Trust “Reposed in the Workmen”.	The Journal of Legal Studies	theory paper		NO
Jacobsen et al., 2018	Why Do We Lie? A Practical Guide to the Dishonesty Literature	Journal of Economic Surveys	overview of the literature		NO
Jakobsen et al., 2017	Making sense of performance regimes: Rebalancing external accountability and internal learning.	Perspectives on Public Management and Governance	theory paper		PA
Jilke and Tummers, 2018	Which Clients are Deserving of Help? A Theoretical Model and Experimental Test	Journal of Public Administration Research and Theory,	deservingness of clients is related to gaming behavior, also mainly theory		PA
Jones, 1991	Ethical Decision Making by Individuals in Organizations: An Issue Contingent Model.	The Academy of Management Review	theory paper		NO
Jones and Euske, 1991	Strategic Misrepresentation in Budgeting	Journal of Public Administration Research and Theory	theory paper		PA
Mass and Van Rinsum, 2013	How control system design influences performance misreporting	Journal of Accounting Research	redundant paper - studies cheating related to peers (Cadsby et al, 2015; Thau et al., 2015)		NO

Meier and Krause, 2003	Conclusion: An Agenda for the Scientific Study of Bureaucracy.	Politics, Policy, and Organizations: Frontiers in the Scientific Study of Bureaucracy	theory paper	PA
Micheli and Neely, 2010	Performance Measurement in the Public Sector in England: Searching for the Golden Thread	Public Administration Review	not an empirical paper	PA
Moore and Gino, 2013	Ethically adrift: How others pull our moral compass from true North, and how we can fix it.	Research in organizational behavior	theory paper	NO
Moynihan, 2009	Through A Glass, Darkly Understanding the Effects of Performance Regimes	Public Performance & Management Review	theory paper	PA
Moynihan, 2010	A Workforce of Cynics? The Effects of Contemporary Reforms on Public Service Motivation	International Public Management Journal	theory paper	PA
Radnor, 2008a	Hitting the target and missing the point? developing an understanding of organizational gaming	Performance Information in the Public Sector	not an empirical paper	PA
Rubin and Hewstone, 1998	Social Identity Theory's Self-Esteem Hypothesis: A Review and Some Suggestions for Clarification.	Personality and Social Psychology Review	theory paper	NO
Siverbo et al., 2019	Conceptualizing dysfunctional consequences of performance measurement in the public sector.	Public Management Review	theory paper	PA
Thau et al., 2015	Unethical for the sake of the group: Risk of social exclusion and pro-group unethical behavior.	Journal of Applied Psychology	redundant paper - studies cheating on behalf of in-group members	NO

			(Cadsby et al, 2015)	
Van Dooren et al., 2012	How to Measure Public Administration Performance	Public Performance & Management Review	not an empirical paper	PA

2.2.2 PA Research: Systems and Organizations

There is a total of 7 articles that deal with system and organizational factors in the public administration literature. Performance gaming is conceptualized as behavior that results in performance data that is desirable but does not meet the intent of the objective or indicator. The conceptualization of performance gaming rests on the idea that it undermines organizational goals usually through some sort of manipulation (Hood, 2006; Moynihan, 2010). Hood (2007, p.100) describes “gaming or cheating as the deliberate massaging or outright fabrication of numbers collected with the intention of improving the position of an individual or organization”. With respect to that definition, Hood describes a gaming typology, with different types of gaming that accounts for the unintended nature of gaming.

Performance systems themselves offer the opportunity for performance gaming behavior because they are often accompanied by features, such as targets, benchmarks, and incentives (Bevan & Hood, 2006). In public organizations, performance measurement and reporting is a means of control. There is a desire for public administrators to be accountable through constant reporting of performance and monitoring of this performance. The assumption, as Smith (1995) states, is that performance systems will yield benefits for the public in efficiency and equity, but this assumption is based on “inadequate modes of production and control” (p. 277), and so a performance system utterly fails if the system itself does not account for its own deficiencies. Performance system target and incentives (Bevan & Hood, 2006; Radnor, 2008b), ensuing competition (Hood, 2006), and performance system design and

measurement flaws (Courty and Marschke, 2007; de Bruijn, 2002) are the primary drivers of performance gaming in the public administration literature.

Organizational culture is important to behavior in a firm (O'Reilly & Chatman, 1996), and so it has been a studied factor for its influence on performance information use (Jennings & Haist, 2004; Moynihan & Landuyt, 2009; Moynihan & Pandey 2010). Taylor (2014) discusses organizational culture as manifesting at different organizational levels to influence performance information use. Hood (2012) theorizes that organizational culture can affect performance system types and thus affect performance systems in perverse ways. Moynihan (2009) explains that performance data use can become embedded in an organization's day-to-day routine and become part of the organization's culture overtime. So, in the public administration literature, organizational culture is an important factor in studying performance use (Andersen and Moynihan, 2016; Cepiku et al., 2017; Destler, 2016). But for performance gaming in particular, researchers are beginning to examine culture because it "has a salient influence on performance information use" (Taylor, 2014, p.9).

2.2.3 Behavioral Insights from Other Disciplines

Other social science disciplines have primarily examined individual and intrinsic factors that lead to performance gaming (Abeler et al., 2014; Ayal et al., 2015; Barkan et al., 2012; Utikal & Fischbacher, 2013). This section focuses on empirical studies from the fields of psychology, business administration, education, and economics, where the primary drivers are self-image, moral identity, self-justification, and other cognitive functions. There are a total of 23 articles that deal with individual factors in the other disciplines examined for this category.

In the literature outside of public administration, manipulation of performance data is studied in the context of various types of unethical behavior, which can encompass more than performance gaming. Unethical behavior describes organizational malfeasance such as “cheating on taxes, insurance fraud, employee theft, academic dishonesty, athletes’ use of illegal drugs, and of course illegal downloading of software and digital content” (Gino et al., 2009, p.393). The way individuals perceive the degree of dishonesty, “the saliency of unethicality,” can dictate whether one engages in the dishonest behavior (Gino et al., 2009, p.393). This is to say that dishonest behavior, if not viewed as dishonest by the participant, can increase the likelihood of it occurring (Baumeister, 1998; Schweitzer & Hsee, 2002). This fact is important for a conceptual understanding of ethical behavior and performance gaming behavior because it makes a parallel that ethicality and gaming can be done unwittingly. Menzel (2005) describes a connection between organizational performance and ethics, describing performance gaming as a type of unethical behavior.

Moral reminders, cues that serve as signals to behave ethically, activate moral identity, a type of identity defined as “the cognitive schema a person holds about his or her moral character” (Aquino et al., 2009, p.124). A study on individuals’ dishonesty revealed that when individuals were given the Ten Commandments before a performance task, individuals acted much more ethical than those not given the commandments (Mazar et al., 2008). The Ten Commandments served as an ethical cue, which activated individuals’ moral identity, prompting moral behavior. People also like to think of themselves as morally honest and “maintaining a favorable self-image is one of the relevant motives” for honest behavior (Fischbacher & Follmi-Heusi, 2013).

Moral identity or moral self-image must be activated in order for it to affect moral behavior (Welsh & Ordóñez, 2014). This is not to say that other factors are not at play, but there is an important role that moral identity plays in moral behavior. Self-concept or moral identity has to be activated or engaged through an activity or a situation (Aquino et al., 2009). People can be primed through a situation or a cue that reminds them of their identity and then this affects their propensity to behave morally (Ayal et al., 2015).

2.3 Analysis of empirical studies

The previous sections discussed performance gaming under three main categories of drivers: system, organizations, and individual drivers. In this section, there will be an in-depth discussion with respect to what has been empirically studied, organized along the three categories of drivers.

2.3.1 System Factors

The system factors that have been studied include types of performance systems and their incentives, pressures and competition effects. And also the design of performance systems, where system design can lead to performance gaming. Courty & Marschke (2007) studied performance gaming in the context of government job training centers. The training centers had a performance measure based on trainee employment status and trainee wages after completion of a job training program. The measurement design was the key factor for gaming behavior because training centers had wide discretion in how they reported successful job training completion and wages. For example, training centers could wait to terminate trainees from the program even if the

trainees were not actively being trained. So, in this study, gaming behavior was operationalized as the accurate reporting of training completion. The practical implications for public management is that designers must regularly detect weaknesses and make adjustments. The public sector has unique aspects for these implications because the authors note that decision-makers at the local level can gain superior understanding of ways to manipulate performance measures, exasperated by the multidimensional and varied nature of objectives in public sector performance measures. Courty & Marschke (2003) developed a theoretical model to show how managers can respond to gaming behavior learned overtime and they apply to this to a case study. Their model is based on system viability theory, which explains that systems are living and evolve and adapt to their environments. Applying this to performance systems involves managers creating monitoring systems, review mechanisms, and then deploying objectives to individual units within the organization, based on information gathered from the monitor and review. When it comes to system factors, the design of a performance system is important to consider as a means of reducing performance gaming.

2.3.2 Organizational Characteristics

Bohte and Meier (2000) is an exemplar study from the organizational thread of performance gaming factors. The context of this study is the school environment, often an ideal setting for studying performance because schools are engaged in more extensive collection and use of data compared to other public sectors. From Bohte & Meier (2000), the factors that contribute to gaming are task demands and resources, which can be externally and internally exerted on an organization. The study focuses on 476 public school districts in Texas. Cheating or gaming behavior is operationalized as the

percentage of students exempted from taking the state standardized test, accounting for legitimate exemptions.

The idea is that schools will want to exclude academically low students from taking the test in order to boost overall standardized test scores. The study uses a theoretical explanation based on goal theory, describing cheating behavior as the act of displacing goals. In goal theory, goals are set as ways to motivate individuals but can become problematic when one goal seems to conflict with another, leading to displacement or substitution of a goal. The practical implications for public management given in the study rest on system design, but also organizational norms are important to mitigate performance gaming.

Organizations must make conscious efforts to reinforce procedures and structures that emphasize goals and appropriate behavior. The practical implications are that there are remedies to reducing cheating via norms and standard, but the specific ways to incorporate these norms into everyday scenarios remains an open question. For organizational factors, individuals are influenced by performance gaming depending on the organization's culture toward morality (Cadsby et al., 2016; Gino et al., 2009) and how that culture manifests itself at different levels (Taylor, 2014).

2.3.3 Individual characteristic

The main driver in the individual category is that of moral identity. "People care about who they are and infer their own values from past choices" (Benabou & Tirole, 2011, p.805). The self-image theme spans across all studies of the individual thread of performance gaming and unethical behavior. It appears to contradict the idea that human

beings are solely economic rationalizers, heavily selfish and void of any moral compass. The individual category of factors focuses on the individual and the cognitive mechanisms that precede decision-making. Two exemplar studies for this thread are that of Schindler & Pfattheicher (2017), who study loss aversion, and also Gino et al. (2013), who study altruistic cheating behavior. In Schindler and Pfattheicher (2017), the researchers conduct two experiments, both consisting of a general population sample, where participants had to roll a die and flip a coin. Cheating behavior is measured by the individuals reporting the number of the die roll or face of the coin flip compared to its probability. In Gino et al. (2013), the authors studied cheating behavior and found that it increases when others can benefit from the cheating. The study consisted of three experiments, each involving college students and tasks that included opportunities to cheat where cheating would knowingly benefit another group other than the cheaters. The researchers found that altruism contributes to positive self-image and allows individuals to justify their dishonest behavior. The practical implications for this study is that beneficiaries are useful for motivating behavior, even when it is dishonest and so individuals must understand how beneficiaries function for a given task or scenario. So, for the category of individual characteristics, moral identity and its influence on moral behavior is the one of the most significant drivers studied (Abeler et al., 2014; Gino et al., 2013)

2.4 Hypotheses

2.4.1 Hypothesis 1: Perceived prosocial impact reduces performance gaming.

Other-oriented constructs, as such as prosociality and public service motivation, offer alternative explanations for how managers make decisions (Belle, 2013b). Because perceived prosocial impact is a nascent area of study (Bolino & Grant, 2016), it is worth examining this particular other-oriented construct. Perceived prosocial impact is defined as the perception that an individual's action at work benefits others (Sonnentag & Grant, 2012). Though it is a burgeoning area of study, research has revealed it has many benefits. Among individual employees, it increases persistence, productivity, and reduces employee errors (Belle, 2013b). It is related to feeling inspired, relaxed, excited, and calm (known collectively as positive affect) (Sonnentag & Grant, 2012). Perceived prosocial impact is further increased when employees are under the direction of transformational leaders (Grant, 2012). Will perceived prosocial impact have these same benefits for performance systems as far as performance gaming is concerned. The hypothesis would be: Perceived prosocial impact is likely to reduce performance gaming. The important element of perceived impact is that individuals must feel their work is beneficial to others. It matters not that the work is actually beneficial, only that the individuals perceive it to be. To measure perceived prosocial impact, a 3-item scale can be used (Grant, 2008b). There are studies that have used perceived prosocial impact as an antecedent to study other behavior (Grant & Campbell 2007; Sonnentag & Grant 2012). These can be examples to follow for public administration's study of performance gaming and perceived impact. In Sonnentag and Grant (2012) participants' perceived

impact is measured at work and participants write in diaries, which are used to gather understanding of how perceived impact influences positive feelings at home.

2.4.2 Hypothesis 2: Public managers who activate moral identity engage in less performance gaming compared to managers who do not.

In the psychology field, there have been studies showing that individuals will lie in order to maintain a favorable image of themselves (Abeler et al., 2014; Fischbacher & Föllmi-Heusi, 2013). When individuals are reminded about their identity, they take on this identity and this affects their level of honesty (Abeler et al., 2014). Cohn et al. (2013) conducted a study on prisoners who were primed about their criminal past and thus had reduced levels of honesty compared to prisoners who were not primed. An important point to make is that individuals think of their moral self in abstract terms and are not likely to engage in moral behavior based solely on their self-concept without a trigger (Ayal et al., 2014). Situational factors trigger the ability of individuals to access their moral identity and thus behave morally (Aquino et al., 2009). The trigger or activation of the moral self is important for self-concept or image to play a significant role in honest behavior. Studies of moral identity as an antecedent of honesty have examined active engagement as necessary in the causal model (Shu et al., 2012). Active engagement can consist of signing one's name on a document before making a report or participating in a morally-laden assignment. So, the hypothesis is extended to state that moral identity must also be activated. The hypothesis would be: public managers who activate moral identity engage in less performance gaming compared to managers who do not activate their moral identity. Moral identity can be measured using a scale by Aquino and Reed (2002) which asks participants to measure their moral self-using a 7-point Likert-type after a list

of moral characteristics and visualizing a person with these traits. Examples of studies can be found in Hertz and Krettenauer's (2016) meta-analysis examining moral identity and moral behavior.

2.4.3 Hypotheses 3 & 4: 3) Perceived prosocial impact and gaming

Verbeke, Volgering, and Hessels (1998, p.313) describe organizational culture as “a system of shared norms and behaviors that are learned by the members of the organization and shape their way of doing”. Because organizational culture shapes behavior, it is always an important internal organizational variable to consider when studying the behavior of individuals within a firm. Scholars have studied how culture improves organizational effectiveness (Schraeder et al., 2005) and the importance of cultural settings being coupled with management practices (Baird & Harrison, 2017; Smith, 1998). Hood (2012) discussed the importance of organizational culture as a category of factor that enhances or obstructs performance. Hood hypothesized that specific types of measurement systems would enhance or obstruct performance, but whether they would obstruct or enhance performance, would depend on the mediating variable of organizational culture. The cultures described were hierarchist, egalitarian and individualist culture. Situational factors trigger the ability of individuals to access their moral identity and thus behave morally (Aquino et al., 2009). Two hypotheses are: the effect of perceived prosocial impact will depend on the organizational culture and also the effect of self-image will depend on culture. Organizational culture can be measured using the Organizational Culture Profile (OCP) (Baird & Harrison, 2017; O'Reilly III et al., 1991; Windsor & Ashkanasy, 1996). The OCP uses five components of culture: Innovation, Teamwork, Outcome Orientation, Respect for People and Attention

to Detail. And it asks respondents to measure the value of each component on a 5-point Likert scale. Andersen & Moynihan (2016) is an example of study that uses organizational culture as an antecedent for performance use in an experimental context. The researchers use a randomized controlled field experiment to study whether public managers are responsive to diversity when they perceive an innovative organizational culture

2.4.4 Experimental research

The public administration field has not benefited, however, from the rich research that can be found in experimental studies, which dominate the literature in other disciplines (Faravelli et al., 2015; Gino et al., 2013; Zhong et al., 2010). Real-life performance schemes can be simulated, where public managers or other individuals are involved in actual performance tasks. The simulated performance task can also include a task where there is opportunity to manipulate performance reports. Faravelli et al. (2015) conducted such an experiment where participants engaged in a performance task and then reported performance that was incentivized. Another innovative approach is an online virtual performance system (Douglas et al., 2019), which is not exactly experimental, but can be altered to run experiments. Douglas et al. (2019) created a virtual public organization where public managers access a role-play tool and report to multiple stakeholders as they apply different performance tools. The researchers study performance management, including gaming behavior from data collected through the tool.

2.5 Conclusion

This review reveals three major categories of drivers that are associated with performance gaming: performance systems factors, individual factors, and organizational factors. This paper makes several hypotheses based on the comparison between public administration and other disciplines. We lack a strong theory for performance gaming. Although public administration scholars have created frameworks and typologies (Courty & Marschke, 2003; Smith, 1995), these frameworks are not strong in their predictive power and are limited in a profound explanation of the phenomenon. They provide implications for management practice, but still lack a clear direction for what variables we should pay attention to and under what conditions. In summary, this paper advocates for research that will lead to theory-building through a study of behavioral factors. This will come from studying many of the individual factors described in this paper (moral identity and perceived prosocial impact). The use of more individual factors and increased experimental designs are the major lessons drawn from the other social science disciplines.

III. ESSAY 2: EXPLAINING EFFORT SUBSTITUTION IN PERFORMANCE SYSTEMS: THE ROLE OF TASK DEMANDS AND MISSION ORIENTATION

3.1 Abstract

This paper contributes to a theory of performance gaming, as it studies why public organizations engage in effort substitution (i.e., directing effort towards rewarded as opposed to unrewarded areas). We argue that effort substitution becomes more likely if tasks are difficult; less likely in the presence of a strong mission orientation; and that mission orientation can mitigate the task demands effect. Examining a five-year panel data set of high schools, we find support for the hypotheses when rewarded and unrewarded measures capture different dimensions of performance. However, results are mixed when rewarded and unrewarded measures are conceptually linked. In more simple terms, this paper suggests that mission-focused organizations are likely to engage in less performance gaming. There is an association between an organization that is aligned to its purpose and mitigating the effects that are linked to performance gaming, such as challenging tasks and limited resources.

3.2 Introduction

Many public organizations are engaging in performance management practices. Performance systems show promise to help organizations improve their efficiency and effectiveness by establishing routines of systematic measurement, reporting, and analysis (Moynihan & Beazley 2016; Van de Walle, 2009; Van Dooren, Bouckaert, & Halligan 2015). At the same time, the link between measurement and improvement is not as explicit as often expected. Organizational factors (Kroll, 2015; Moynihan & Pandey, 2010) and individual biases (Nielsen & Moynihan, 2017; Olsen, 2015) constrain the use of performance information for decision-making, and purposeful use may not necessarily create immediate improvements in organizational and societal outcomes (Gerrish, 2016; Kroll, 2017).

Meanwhile, evidence is accumulating that dysfunctions of performance systems can harm clients and citizens. Specifically, data gaming – which we conceptualize as generating positive performance data without achieving the actual objective behind the indicator – can be detrimental. Examples of performance gaming include the downgrading of crime to mask bad statistics, “cream skimming” of easy-to-serve clients in job training programs, or a tendency to teach-to-the-test in the school system (Eterno & Silverman, 2012; Heinrich & Marschke, 2010; Radin, 2006). Although we do not believe that gaming and cheating behaviors will dominate every performance system (and our thinking is informed by Kelman & Friedman, 2009; Kerpershoek et al., 2016 as well as Wenger et al., 2008), we note that such observations have occurred too frequently across several contexts to be dismissed as exceptions or outliers, and we suggest that

most systems are likely to experience some level of gaming (e.g., Bohte & Meier, 2000; Hood 2006; Jacob & Levitt, 2003; Soss, Fording, & Schram 2011).

Despite the overwhelming evidence suggesting that gaming is a real threat to most performance systems, we lack organizational theory explaining why gaming is more likely to occur in some cases than in others. In fact, most theories focus on the system itself as the explanatory variable arguing, for example, that static regimes with high-powered incentives and a great deal of social pressure are very likely to experience performance gaming (Eterno & Silverman, 2012; Heinrich & Marschke, 2010). While we agree that system configuration shapes behavior, we also think there is more to learn about the differences in the behavior of organizations that are all subjected to the same performance system. That is, once we hold the incentive structure constant, how can we explain why some organizations are more likely to engage in gaming behavior than others?

This paper contributes to such a theory of performance gaming. Specifically, we study one type of gaming known as effort substitution, where organizations tend to focus on rewarded areas at the expense of unrewarded ones. We argue that effort substitution is a function of task demands, mission orientation, and the interactions of these two factors. In cases in which tasks are more difficult, organizations have a hard time meeting performance targets and, hence, may shift attention and resources from unrewarded to rewarded areas. At the same time, a strong mission orientation may prevent organizations from trading off multiple purposes to the point, where mission orientation mitigates the effect of task demands on effort substitution.

We examine these hypotheses in the context of schools in the United States, and we do so for two reasons. First, the school system shares important characteristics with public organizations and agencies more broadly and, hence, constitutes a fairly representative case (O'Toole & Meier, 2006; 2011). Second, schools in the US are a particularly promising case to study performance management systems because schools (specifically the K-12 sector) have been exposed to more than a decade of standardized testing and accountability mechanisms, and this policy area has also seen its fair share of performance gaming (Jacob & Levitt, 2003; Lewis & Triantafillou, 2012).

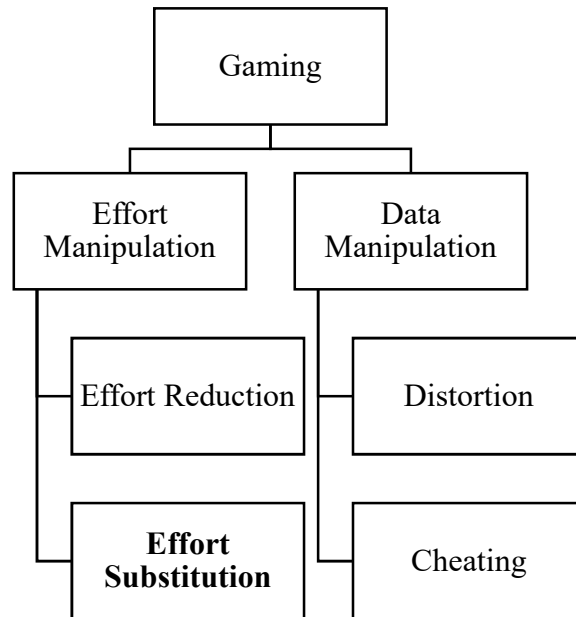
In particular, we employ a five-year panel data set of 64 high schools in one school district, thereby holding constant the performance regime under which all the schools operate. We capture effort substitution by creating variables that put schools' performance on rewarded measures in relation to their performance on unrewarded ones. One effort substitution variable is based on rewarded and unrewarded measures that are conceptually linked, while the other is based on measures that are minimally connected. We find support for the proposed effects of task demands, mission orientation, and their interaction, however, mostly in cases in which rewarded and unrewarded measures pick up on different objectives and offer real choices. Our findings are much more mixed if rewarded and unrewarded measures are conceptually linked.

3.3 Gaming of Performance Systems

In this section, we review and organize the literature on performance gaming, and we explain how our work on effort substitution fits with previous scholarship. In our

typology, we differentiate between two types of gaming – the manipulation of data and effort (see Figure 3).

Figure 3: Gaming Typology



Data manipulation is the fudging of numbers, where performance improves on paper, but without any actual increases in output. Kelman and Friedman (2009, p. 924) describe this as simply gaming or the consumption of real resources “with no genuine performance improvement even on the measured indicator.” Smith (1995) introduces two terms: misrepresentation, where data is misrepresented or distorted, and ossification, where a data indicator is intentionally kept in the performance system though it is no longer needed or relevant, thus causing performance to be skewed. Bohte and Meier (2000) describe data manipulation as a form of organizational cheating. All these terms fall under the umbrella term of data manipulation because each term deals with altering performance information without producing actual output or improvement. Distortion and cheating are the two main forms of data manipulation (Hood, 2006). The distortion of data involves creatively interpreting, reporting, or rearranging data to make it favourable.

In contrast, cheating behavior is lying or complete fabrication of data and is the riskiest form of gaming, as it can carry serious consequences if discovered.

The public administration literature on data manipulation has focused mainly on the distortion of data. Courty and Marschke (2007) study data distortion and show how managers manipulated program termination dates of trainees so they could terminate trainees when their wages and employment status had reached acceptable performance levels. The study reveals that managers can game the system by learning it. Another example of creative distortion comes from Bohte and Meier (2000), who examine how superintendents of school districts in Texas exempted certain students from taking standardized tests. Performance numbers went up without any actual improvement in student achievement. Bevan and Hood (2006) find that response times for ambulances had been “corrected” to the target time without explanation, which is yet another example of gaming behavior.

The second major form of gaming is effort manipulation, which describes ways in which managers manipulate effort or resources to game the system. Under effort manipulation, there exists the sub-categories of effort substitution and effort reduction. Effort reduction, described by Smith (1995) as sub-optimization, involves reducing effort without necessarily transferring the effort elsewhere. Effort reduction is also described using the terms ratchet effect and threshold effect (Bevan & Hood, 2006). The ratchet effect is behavior, which occurs when managers know that they must meet or exceed last year’s numbers. In effect, they reduce effort this year so they can more easily reach the target next year. The threshold effect occurs when a particular threshold must be reached by everyone in the organization. In essence, low performers will be pressured to reach the

threshold, while high performers, who usually surpass the threshold, will deteriorate their performance to drop back to the threshold and, hence, willingly lower their performance.

Effort substitution occurs when more effort is given to a measured indicator at the expense of an unmeasured indicator – basically effort is substituted for the unmeasured indicator (Kelman & Friedman, 2009). Other scholars have labelled this phenomenon goal displacement (Bohte & Meier, 2000) or output distortion (Hood, 2006). An example of this, in the context of public schools, would be to reduce instruction in social studies (since it is not assessed on most standardized tests) and to increase instruction in reading and math (which are assessed on standardized tests). Effort substitution results in actual improvement on the measured indicator, but it usually comes at the expense of the unmeasured indicator. Cream skimming may fall under effort substitution if it involves serving the most favourable clients, while neglecting less favourable clients (Koning & Heinrich, 2013). Such behavior will involve effort substitution, as resources will be taken away from the unfavourable clients in order to support efforts for the more favourable clients.

Effort substitution is different from making purposeful strategic choices. In fact, prioritization is legitimate and often the result from strategic planning efforts. For example, in some situations, public organizations may opt to prioritize quality over efficiency, and in other cases, they may make the reversed choice. However, if an organization makes the choice to focus on quality (as opposed to efficiency, for example), we expect it – if considered successful – to do well on multiple measures of quality, rewarded and unrewarded ones. However, if the organization does much better on rewarded than unrewarded indicators, we consider this evidence for effort substitution.

There have been a handful of studies on effort substitution. For example, Bevan and Hood (2006) studied storming, a form of effort substitution where employees storm a task by sucking up effort from other departments in order to meet an impending evaluation date. This was observed in their study of the U.K. Commission of Health where hospital workers cancelled operations and drafted in other medical staff during the days leading up to an evaluation. Koning and Heinrich (2013) observe effort substitution in their study of workers who received disability insurance versus workers who received unemployment insurance. The disability insured were harder to serve and as result, they had longer job searches – evidence that effort had been reduced in serving the harder-to-place disability insured workers.

This paper contributes to the literature on effort substitution, which is one of four types of performance gaming. Our understanding of effort substitution is descriptive, not normative. If organizations focus on objective A but only do well on the related rewarded indicators (e.g., A1 and A2) but not the unrewarded ones (e.g., A3 and A4), we consider this effort substitution. While we observe this behavioral pattern, we make no normative claims about who is at fault – individuals who make choices, interest groups creating pressure, or the performance system that generates constraints. We are interested in documenting behavioral outcomes and understanding related organizational triggers, but our study does not aim at disentangling multiple individuals' intentions, actions, and motivations. While the latter could be certainly of interest to readers, it would also require conducting a different empirical study. In the following section, we outline a theory of effort substitution, which we then test in the subsequent sections.

3.4 Task Demands and Mission Orientation

Our theoretical arguments are based on a set of assumptions about performance systems. First, in the creation of performance management systems, elected officials are the ultimate designers of performance systems and these regulators have a clear intention of the performance they seek to stimulate. The relationship of the regulator (elected officials) and the agents (bureaucrats) is explained by principal-agent theory which describes agents as being controlled and monitored by elected officials (the regulators or principals) via the collection of performance data (Moynihan, 2008). Without a performance system, the agents are likely to shirk the assigned task and act opportunistically. Second, while some bureaucrats are driven by self-interest, others may value prosocial motives and act as stewards of the organizational mission. Organizations can shape mission orientation through structure and culture, and a mission-based culture is seen as a key feature of effective government and provides a framework for goal-setting, fostering motivation and results (Rainey & Steinbauer, 1999; Wolf, 1993). Hence, we make the assumption that performance systems contend with varying levels of mission-orientation. Our third assumption is that performance systems are multifaceted and thus agents are multitasking to achieve proficiency for various performance indicators. The performance system must contend with these different areas of performance for the same organization (Van Dooren et al., 2015). This is a peril of performance systems because agents may choose to attend to some goals while neglecting other goals or completely avoiding the objectives (Moynihan, 2008; Smith, 1995).

Many performance systems are being set up for the purpose of benchmarking and comparing agencies or units that conduct similar tasks (Gerrish, 2016; Van Dooren et al., 2015). Examples include schools, police departments, or hospitals (Eterno & Silverman, 2012; Jacob & Levitt, 2003; Kelman & Friedman, 2009). A main feature of such systems is that they apply the same indicators and targets to assess agencies, rank them, and reward or sanction based on performance. However, while the task these agencies perform may be identical, the level of difficulty is likely to vary. Task demands are shaped by the conditions under which the task is performed (Robinson, 2001, p. 32f.). In a government setting, the needs of minority and low-income clients often make the service provision more demanding (O'Toole & Meier, 2006, p. 102). Examples of clients who may be harder to serve include schools where students come to class with issues that impede learning, or hospitals where patients have complex healthcare needs.

In cases in which tasks are difficult but performance rewards or sanctions substantial, agencies will likely be punished for poor performance, despite the fact that these agencies must cope with bigger constraints. If said agencies do not deem it feasible to enhance measured performance due to actual output improvements, they might engage in other practices such as gaming. One particular strategy to deal with performance pressure is to shift effort, attention, and resources to measured activities and objectives at the expense of other important – but unmeasured – tasks (Bohte & Meier, 2000; Koning & Heinrich, 2013). While some systems are able to attenuate the negative task demands effect through a focus on relative gains, most systems assign at least some importance to absolute performance levels or minimum targets, and only a few systems create benchmarking clusters based on similar task demands.

Agencies confronted with daunting tasks demands may choose to engage in effort substitution because such practices could be considered somewhat legitimate in the face of rigid performance pressure, and they do not require any data manipulation. Effort substitution is also feasible because it exploits the professional discretion that most public agencies have. Typically, political authorities identify difficult problems and then bureaucrats are tasked with solving these complex problems, often with limited resources. A dilemma occurs because bureaucrats are given only vague guidance as to how they should solve these complex problems (Goodsell, 1994). Due to broad discretion in policy-making, bureaucracies may opt to “creative” ways to solve problems, subverting performance system mechanisms in ways that undermine organizational goals (Soss et al., 2011).

As argued above, effort substitution can have very detrimental effects on agency clients and outcomes. In fact, dealing with difficult tasks in a government setting often means working with underserved populations, which may end up doubly disadvantaged if improvements in one area are traded for null effects in others. Considering the literature on performance pressure, varying task demands, and effort substitution we summarized in this section, we hypothesize the following:

H₁: Task demands will increase effort substitution.

An organizational mission articulates the purpose of the organization: “why do we exist?” In a public or non-profit context, organizational missions are mostly about societal impact, making a difference, or helping clients – hence, the purpose is prosocial (Knies & Leisink, 2018; Moore, 1995). At the same time, mission statements are

inclusive and try to be broad enough to serve as “umbrellas” covering an organization’s many functions and activities. To this extent, missions are (at least partially) at odds with performance metrics since the latter are the result of strategic prioritization and the identification of key performance areas. But even in cases in which performance priorities and mission overlap, a set of quantitative indicators is rarely able to comprehensively capture the all-encompassing abstract of what is truly desired (Binning & Lebreton, 2009).

Due to these conceptual and empirical tensions between a mission and performance metrics, organizations with a strong mission orientation may pay less attention to a specific set of performance indicators, even when put under pressure. Said organizations are likely to consider that performance metrics provide an incomplete picture of the true organizational purpose. Hence, organizations with a strong mission orientation will not give in to performance pressure by shifting their efforts away from unmeasured to measured tasks and objectives. Research on individuals and teams who have a strong sense of purpose and are prone to prosocial motives supports this notion. Said individuals are less likely to engage in gaming and, as Korsgaard et al., (1997) find, more likely to appreciate performance feedback – even if negative – and act upon it in order to improve. Similarly, studies show that prosocial motives foster positive and voluntary work behaviors, subsumed under the label of organizational citizenship, including team behaviors (Hu & Liden, 2015) and behaviors that benefit individuals outside of the organizations (Takeuchi et al., 2015).

A strong mission orientation can also help to set mission-related norms that affect everyday behavior. As argued before, missions of public and non-profit organizations

tend to be prosocial and, hence, support prosocial norms and behaviors. Experimental research in behavioral economics shows that social norms are likely to influence cheating behavior. For example, being asked to recall the Ten Commandments before conducting an incentivized performance task made cheating less likely (Ayal et al., 2015). Further, cheating becomes more likely if similar others (in-group members) indicate that they are cheating when performing the task (Gino et al., 2009). Overall, this indicates that a strong orientation on mission may prevent effort substitution as mission focus emphasizes prosocial norms and behaviors over financial rewards.

In addition to the independent effect of mission orientation, this variable could also interact with task demands. That is, mission orientation could buffer the reinforcing effects of task demands on effort substitution.

H₂: Mission orientation decreases effort substitution.

H₃: Mission orientation mitigates the effect of task demands on effort substitution.

3.5 Data and Methods

3.5.1 Research Context and Data

We examine our theory in the context of schools and the school system. We do this for two reasons. First, the school system is representative of many cases within public administration more broadly. It is a hierarchical bureaucracy consisting of multiple tiers (states, districts, and schools), while at the same time giving a great deal of discretion to highly professionalized actors at the different tiers, including street-level bureaucrats (the teachers in the classrooms). Schools are vastly diverse, and comparisons

across schools offer a large amount of variation in important variables, such as task demands, which is one of the main factors in our theory. School districts have been labelled “the most common public organization in the United States” (Meier & O’Toole, 2009, p. 7), and the fact that virtually every country has publicly-managed school systems makes them a fairly representative case.

Second, schools are a particularly promising case to study performance management systems. The area of education (particularly schools) has a longstanding experience with performance systems. Most schools and school districts have gone through at least one decade of using standardized testing, performance comparisons, and accountability mechanisms, turning this policy area into a “fertile ground for demonstrating changing performance measurement regimes and their consequences” (Lewis & Triantafillou, 2012, p. 606). Furthermore, many cases of the dysfunctions of performance systems (gaming and cheating) have been reported in the area of education and schools (for a large-scale analysis on the matter, see Jacob & Levitt, 2003). This includes prominent cases of teachers and school administrators erasing and correcting answers on the high-stakes state standardized tests as well as cases of falsifying graduation rates (e.g., Kasperkevic, 2015; Strauss, 2018).

To test our hypotheses, we study organizations (schools) within one performance regime (the school district’s system). By focusing on one school district, we aim at separating performance system-effects on gaming from organization-level variables because the former are held constant and, hence, equally present for all schools. In

particular, our sample consists of 268 observations from 64 high schools¹ in the Miami-Dade County Public Schools (M-DCPS) district for the 2012-2016 school years. We chose this case because of the district's longstanding experience with performance metrics predating the state system, its relevance within Florida due to population size, and the availability of data for all relevant variables for multiple years. Further, Dade County is often considered a "laboratory" for the nation because the population of the United States is projected to become similarly diverse to that of the county.

While all school districts in Florida apply the same state-wide performance metrics, the implementation of the system and components to incentivize performance can vary among districts. The state-wide system assigns school grades between A and F to schools based on 11 performance criteria related to achievement, learning gains, graduation rate, and acceleration success (Florida Department of Education, 2017). The Florida School Recognition Program is a state-wide program that provides financial rewards to schools and individual educators based on the school grades, therefore incentivizing the 11 performance measures. In addition to financial rewards, school grades also serve as measures by which schools are placed under additional scrutiny both by the individual school district and the state.

Most of the performance-related funding comes from the State of Florida. Schools that improve a letter grade, maintain an improved letter grade from the previous year, or receive a "commendable rating" are rewarded with a cash performance bonus from the

¹ We included all high schools that are a part of the district's performance metrics, which excludes special needs schools and schools for juvenile delinquents. All data used for our analysis are publicly available: <http://oada.dadeschools.net/SchoolPerformanceData/SchoolPerformanceData.asp> as well as <http://drs.dadeschools.net/SchoolClimateSurvey/SCS.asp> (both retrieved on November 2, 2017).

Florida School Recognition Program. This includes rewards at the individual level for principals, teachers, and staff (Florida Department of Education, 2018). However, while the state pulls the levers in rewards and sanctions through funding, it is the local districts that actually implement the actions through their discretionary use of funds (Iasevoli, 2016) and their restructuring of individual schools, which include transferring administrators and teachers when schools do not meet standards (Emma, 2015; Solochek, 2018). Local school districts also reward and sanction schools through promotions of successful administrators and teachers and demotions of those deemed unsatisfactory (Cohen et al. 2012; Travis, 2016). Further, school districts have a great deal of discretion when it comes to the use of funds from block grants to incentivize performance, such as the Teacher Incentive Fund from the U.S. Department of Education.

Florida has several lists of sanctioned schools. One list is labelled the “300 Lowest Performing Elementary Schools” and another list is labelled “Persistently Low-Performing Schools.” Though such schools receive additional resources and support to resolve performance inadequacies, the stigma that comes with being on such a list undoubtedly hurts a school’s image. To add to the image problem, the pressure from intense oversight on individual teachers and school-level leadership can be overwhelming. Studies have shown that school staff turnover is related to such external pressures (e.g., Boyd et al., 2005).

Furthermore, the media often prints the name of school grades, so the public is fully aware of a school’s status. Neighborhoods and even whole local municipalities will tout the success of local schools in order to attract homeowners and even businesses to an area. This is not just an occurrence in Florida, but across the nation. An article in the

Washington Post summed it up best by stating that good schools are a top factor for homeowners nationwide when deciding on a home to buy (Lerner, 2015). Because of the significance of school accountability grades and the salience of the issue, it is a high likelihood that school officials will pay particular attention to these 11 incentivized components that make up school accountability grades.

3.5.2 Measuring Effort Substitution

To capture effort substitution, we create variables that measure discrepancies between rewarded and unrewarded school performance metrics. One rewarded measure we employ is the graduation rate, whereas unrewarded – but important – indicators are the dropout, attendance, and mobility (percent of students who transferred in or out a given school) rates. While we explain in more detail how we created our effort substitution measures below, this is the approach in a nutshell: We subtract the unrewarded indicators (dropout, attendance, and mobility) from the rewarded indicator (graduation rate) and consider high scores as evidence of effort substitution. If schools do not substitute effort, then scores of rewarded and unrewarded metrics should roughly be of equal magnitude, resulting in a net difference close to zero. If, however, schools are doing considerably better on rewarded as opposed to unrewarded statistics, then the net difference will be positive, suggesting the existence of effort substitution.

In order to provide a comprehensive test of our hypotheses, we replicate our analysis in two contexts – one in which rewarded and unrewarded measures are conceptually linked and one in which they are not. In line with prior work, we expect the graduation rate (our rewarded measure) to be at least somewhat conceptually tied to our

three unrewarded measures (dropout, attendance, and mobility) (Mishel & Roy, 2006; O'Toole & Meier, 2011). Although it seems possible to improve a school's graduation rate, while neglecting dropout, attendance, and mobility, at some point improvements in graduation will only be feasible by also enhancing at least one or two of the three unrewarded metrics. Hence, we introduce a second rewarded measure, which we believe is less related to the three metrics that are not rewarded.

This measure is the acceleration success rate, which is the “percentage of graduates from the graduation rate cohort who earned a passing score on an acceleration examination, such as advanced placement or the International Baccalaureate exam, a passing grade in a dual enrolment course that qualified for college credit, or earned an industry certification” (Florida Department of Education 2016, p.2). In essence, the acceleration success rate is a measure of the percentage of students who did not just graduate, but obtained college or career readiness qualifications. Since this is a very high standard of performance, which is only relevant for a subpopulation of students, we believe schools are able to improve in this area, while entirely neglecting overall dropout, attendance, or mobility rates.

In total, we are using two rewarded and three unrewarded measures of performance, which are listed in Table 3 and accompanied by brief definitions. Our rewarded measures (graduation and acceleration) are cohort-based, which is why we also converted the dropout rate into a cohort-based variable. Attendance and mobility rates could not be transformed into cohort-based measures because both metrics were not broken down by grade level for each year in the data set. However, both metrics did not show too much fluctuations within schools over the panel years. Before subtracting

unrewarded from rewarded measures, we created z-scores to make all five scales comparable. The scales for dropout and mobility were reversed, so that high z-scores on all five measures capture “positive” performance.

Table 3: Effort Substitution Measures and Factor Structure

Measures	Definition	Factor Loadings ^c	
		Acceleration Gaming	Graduation Gaming
Graduation vs. Dropout	This is the difference between graduation rates ^a (rewarded) and (reversed coded) dropout rates ^a (unrewarded).	-0.070	0.927
Graduation vs. Attendance	This is the difference between graduation rates ^a (rewarded) and attendance rates ^b (unrewarded).	0.138	0.784
Graduation vs. Mobility	This is the difference between graduation rates ^a (rewarded) and (reversed coded) mobility rates ^b (unrewarded).	-0.032	0.918
Acceleration vs. Dropout	This is the difference between acceleration success rates ^a (rewarded) and (reversed coded) dropout rates ^a (unrewarded).	0.726	0.253
Acceleration vs. Attendance	This is the difference between acceleration success rates ^a (rewarded) and attendance rates ^b (unrewarded).	0.995	-0.093
Acceleration vs. Mobility	This is the difference between acceleration success rates ^a (rewarded) and (reversed coded) mobility rates ^b (unrewarded).	0.950	-0.016
Eigenvalue		3.715	1.206

Notes: a = cohort-based; b = single-year based; c = principal component factoring with promax rotation; acceleration success rate = percentage of students acquiring college credits or industrial credits for trade school prior to graduation; mobility rate = percent of students who transferred in or out a given school.

The combination of all rewarded and unrewarded measures allows us to create six effort-substitution variables. For the purpose of data reduction, we factor analysed the six variables, and the results can also be found in Table 3. The six variables load on two factors – acceleration gaming and graduation gaming – with significant Eigenvalues and no considerable cross-loadings. We use these two factor scores as dependent variables in our effort substitution regression models.

3.5.3 Independent Variables

One of our main independent variables is task demands, measured as the percentage of Black students in a school and also socioeconomic status, based on the percentage of students on free or reduced lunch. In the context of schools, the minority status of students is often a proxy measure of task demands because schools with large minority populations tend to have lower standardized test scores (Bothe & Meier, 2000; Rong & Grant, 1992).²

The second variable of main interest is mission orientation. We distinguish levels of mission orientation using a dummy variable for public versus charter schools, assuming that – on average – charter schools are more mission oriented. Charter schools are often founded for a specific purpose, and this purpose is articulated in their charter and mission statement, guiding their everyday operations. Although we expect variation across charter schools, research has shown that they are also somewhat homogenous

² Hispanic students are also often considered to be a disenfranchised group that requires higher teaching demands. However, unlike in Texas or California, Hispanic students do not have the same status in South Florida (Gay, 2006; Mohl, 1990). In fact, in our data set schools' percentage of Hispanic students is negatively correlated with the percentage of Black and low-income students, which is why we do not include this group in our measures of task demands.

regarding a strong internal focus on the implementation of their charter/mission (Ford & Ihrke, 2016) as well as a strong teacher community (a construct that includes teachers' shared beliefs and values) (Cannata, 2007). Since charter schools may not just differ from public schools regarding mission and purpose but also with respect to performance, size, climate, and resources, we control for these factors, whose measurement we explain further below.

Of course, some public schools may be more mission oriented than some charter schools, which will likely increase measurement error in this dummy variable and noise in our statistical findings. Hence, we employ a second measure that we call "mission specificity." To create this measure, we conducted a content analysis of all schools' mission statements (every school has such a statement), similar to approaches used in previous scholarship (Boerema, 2006; Garrow & Gursky 2013; see also Desmidt, 2016). We employ a dummy variable to differentiate between "any mission" (0) and missions that emphasize a philosophical value, such as having a particular worldview or a non-traditional service delivery approach (1). The idea is that schools with missions, which champion a specific educational approach, may also provide specific operational guidance to educators and, thus, are more influential in shaping behavior. Administrators are less likely to ignore a mission if it reflects the unique purpose of the school's existence as opposed to any school's generic purpose, which is simply to educate students.

We corroborated the mission information on the schools by also reviewing websites or school improvement plans for further evidence on their approach to education. The following statement is an example for a generic mission:

We provide the highest quality education so that all of our students are empowered to lead productive and fulfilling lives as lifelong learners and responsible citizens.

The targeting of a specific student group (boys versus girls or a focus on the arts, sciences, or college readiness) was not sufficient to be coded “1” on our mission specificity variables, which required a unique philosophical or educational approach. The following statements are examples of missions we coded “1.”

[...] is dedicated to empower students with the knowledge to understand and care for the environment. [...] will provide students with preparation, encouragement and inspiration for higher learning. Our ultimate goal is to prepare students to become environmental ambassadors of humanity and to deliver the message that a sustainable relationship with the Earth is possible.

[...] The core philosophy and underlying purpose of [...] is reflected in the following concepts: High expectations for students and teachers: Creative endeavours as an integral part of the growth and development of all students, Character development, Increase of self-esteem through mechanisms that ensure the improvement of students’ self-image as learners, Parental involvement, Student ownership of their learning through self-awareness of their learning styles and self-monitoring of their learning, Student and teacher accountability, Commitment to the ideal that success breeds success [...]

At [...], our faculty is committed to empowering our students through mentorship to be held accountable by teaching them to embrace responsibility, demonstrate mutual respect, and engage in open communication. Our continuous collaboration of all stakeholders will provide a safe and nurturing environment which promotes students’ social-emotional and academic growth. [...] Through our endeavours and dedication to community service, our students will achieve their full potential and become productive members of society.

There was no change over time in mission statements in our study period, which is why this variable is static within schools. The fact that our charter-school and mission-specificity dummy variables are positively but moderately correlated ($r=0.38$) shows that both pick up on the same latent idea (“mission orientation”) but are still different enough to be captured by two variables.

We include the following control variables, all measured at the school level: performance grade, size, climate, focus on performance data, and lack of resources. We do this for two reasons. First, we want to account for differences between public and charter schools other than mission orientation. Second, we think these variables may impact gaming behavior. For example, gaming may be more restricted in larger schools that have more structure and formalization and involve more actors in the decision-making. At the same time, a lack of resources could make administrators more desperate and likely to engage in practices such as effort substitution.

School size is measured by the number of students. The other three variables are constructed based on single items from the M-DCPS annual school climate survey given to staff members and averaged at each school. The three variables are measured using 5-point Likert scales, and their wording is as follows: “The overall climate or atmosphere at my school is positive and helps students learn” (climate); “annual teacher evaluations are used to improve teacher performance” (performance focus); and “I am limited by insufficient resources (e.g., books, equipment, supplies etc.)” (lack of resources). The descriptives and correlations of all variables can be found in Table 4.

Table 4: Descriptive Statistics and Correlations

	Mean (SD)	Range	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. Acceleration gaming	0.00 (1.00)	-4.47 – 2.40	1.00									
2. Graduation gaming	0.00 (1.00)	-9.60 – 2.08	0.48	1.00								
3. Percent low-income students	69.47 (19.40)	4 – 96	0.24	-	1.00							
4. Percent Black students	23.29 (28.18)	0 – 95.5	0.35	0.18	0.35	1.00						
5. Charter Schools	0.23 (--)	0 – 1	-	-	-	-	1.00					
6. Mission specificity	0.17 (--)	0 – 1	0.51	0.05	0.22	0.32		1.00				
7. Performance grade	3.23 (0.88)	1 – 4	-	0.03	-	-	0.15	0.13	1.00			
8. Size	1,688.25 (1,014.04)	112 – 4,244	0.06		0.07	0.06						
9. Climate	4.05 (0.44)	2.6 – 5	0.23	-	0.21	-	-	-	-	1.00		
10. Focus on performance data	3.41 (0.38)	2.26 – 4.82	-	-	-	-	0.35	0.13	0.07	-	1.00	
11. Lack of resources	2.85 (0.63)	1.21 – 4.68	0.21	0.04	0.28	0.36				0.23		
			0.32	0.09	0.13	0.34	0.47	0.15	0.06	-	0.38	1.00
			0.14	0.01	0.11	0.36	-	-	0.09	0.16	-	-
							0.23	0.06			0.34	0.48

Table 5: Acceleration gaming models: Unlinked performance measures

	DV: Focus on college and career acceleration as opposed to unrewarded measures [^]			
Models	(1)	(2)	(3)	(4)
<u>Task demands</u>				
Percent low-income students	0.007** (0.004)	0.005 (0.181)	0.007** (0.020)	0.003 (0.429)
Percent Black students	0.006** (0.016)	0.011** (0.000)	0.006** (0.008)	0.012** (0.000)
<u>Mission orientation</u>				
Schools with charters	-1.381** (0.000)		-2.450** (0.000)	
Mission specificity		-0.279* (0.082)		-0.371** (0.014)
<u>Interactions</u>				
Percent low-income students x schools with charters			-0.008 (0.142)	
Percent black students x schools with charters			-0.059** (0.000)	
Percent low-income students x mission specificity				0.008 (0.178)
Percent black students x mission specificity				-0.011* (0.051)
<u>Control variables</u>				
Performance grade	0.084* (0.085)	0.057 (0.396)	0.089* (0.066)	0.068 (0.313)
Size	-0.000** (0.021)	0.000 (0.266)	-0.000** (0.027)	0.000 (0.223)
Climate	0.119 (0.380)	0.005 (0.973)	0.087 (0.507)	-0.023 (0.881)
Focus on performance data	-0.178 (0.172)	-0.536** (0.000)	-0.202 (0.105)	-0.517** (0.000)
Lack of resources	-0.071 (0.535)	-0.147 (0.240)	-0.047 (0.692)	-0.146 (0.248)
Constant	-0.044 (0.960)	1.293 (0.258)	0.712 (0.415)	1.898* (0.069)
Dummies for years 2012-2016 included	Yes	Yes	Yes	Yes
	N	268	268	268
	R ²	0.425	0.259	0.449
	adj. R ²	0.398	0.224	0.418

Notes: * p < 0.1, ** p < 0.05; p-values in parentheses; robust standard errors clustered at the school level; pooled OLS regressions; ^ unrewarded measures: dropout, attendance, and mobility rates

Table 6: Graduation gaming models: Linked performance measures

	DV: Focus on graduation rates as opposed to unrewarded measures [^]			
Models	(1)	(2)	(3)	(4)
<u>Task difficulty</u>				
Percent low-income students	-0.007** (0.034)	-0.007** (0.022)	-0.013** (0.000)	-0.009** (0.010)
Percent Black students	0.008** (0.010)	0.008** (0.001)	0.010** (0.001)	0.010** (0.000)
<u>Mission orientation</u>				
Schools with charters	-0.129 (0.694)		-1.108** (0.043)	
Mission specificity		0.007 (0.973)		-0.113 (0.475)
<u>Interactions</u>				
Percent low-income students x schools with charters			0.010 (0.151)	
Percent black students x schools with charters			-0.061** (0.014)	
Percent low-income students x mission specificity				0.009 (0.199)
Percent black students x mission specificity				-0.014** (0.010)
<u>Controls</u>				
Performance grade	0.049 (0.309)	0.044 (0.366)	0.066 (0.162)	0.057 (0.263)
Size	-0.000 (0.562)	-0.000 (0.591)	-0.000 (0.818)	-0.000 (0.725)
Climate	-0.040 (0.874)	-0.053 (0.822)	-0.088 (0.719)	-0.088 (0.711)
Focus on performance data	-0.169 (0.265)	-0.205 (0.139)	-0.134 (0.302)	-0.185 (0.182)
Lack of resources	-0.179 (0.333)	-0.187 (0.283)	-0.088 (0.647)	-0.190 (0.289)
Constant	1.469 (0.301)	1.618 (0.175)	0.837 (0.549)	1.328 (0.278)
Dummies for years 2012-2016 included	Yes	Yes	Yes	Yes
	N	268	268	268
	R ²	0.088	0.086	0.140
	adj. R ²	0.045	0.043	0.093
				0.052

Notes: * p < 0.1, ** p < 0.05; p-values in parentheses; robust standard errors clustered at the school level; pooled OLS regressions; ^ unrewarded measures: dropout, attendance, and mobility rates

3.6 Results

To test our hypotheses, we run models using our two effort substitution measures as dependent variables (acceleration and graduation gaming) and our measures of task demands, mission orientation, and their interaction as well as a set of controls as independent variables.³ All continuous measures that make up the interaction terms were mean-centred before we multiplied them to simplify the interpretation of the interactive models. We pool all observations across years and estimate OLS regressions since there is no or very little change across time in most of our independent variables of interest, which conceptually is in line with previous research on schools and other public organizations (Andrews, 2010; O'Toole & Meier, 2010 and 2011) and statistically the most appropriate choice for our data.⁴ Positive regression coefficients indicate increases in effort substitution in favour of the rewarded performance measure.

Table 5 shows the regression results using acceleration gaming (a focus on college and career acceleration as opposed to dropout, attendance, and mobility rates) as the dependent variable. Here, rewarded (acceleration) and unrewarded (the other three rates) performance measures are not strongly conceptually tied to each other, which is

³ Our main specification is this: $ES = \beta_0 + \beta_1 TD + \beta_2 MO + \beta_3 (TD \times MO) + \beta_4 C + \mu + \varepsilon$, where all betas are estimable parameters, μ represents the year fixed effects, and ε is the error term. We run multiple variations of this equation because effort substitution (ES) is captured through graduation and acceleration gaming; task demands (TD) through percent low-income students and percent Black students; and mission orientation (MO) through schools with charters and mission specificity. The controls (C) in all models are performance grade, size, climate, focus on performance data, and lack of resources. OLS Diagnostics are found under Appendix 3.

⁴ The use of school-fixed effects would require substantive over-time change in our main variables, and the Hausman Test rejects the application of random effects in at least half of our models. To account for the panel structure of the data, we employ robust standard errors clustered as the school level and a set of year dummies.

why it is possible to improve the former and neglect the latter. Models one and two examine the effects of task demands and mission orientation, where the first model employs the charter school dummy and the second model the mission specificity dummy to capture mission orientation. Model three and four test the interactive effects, where the former includes the interactions with the charter dummy and the latter those with the specificity dummy.

Models one and two provide evidence for the effect of task demands and mission orientation. All six coefficients show the expected signs and five out of these six are statistically significant. Hence, task demands seem to reinforce effort substitution, and mission orientation reduces it. In the interactions models three and four, we see that that three out of four coefficients show the expected negative sign and two coefficients (the interactions with percent Black students) achieve statistical significance. This is at least partial support for the hypothesis that mission orientation can attenuate the reinforcing effect of task demands on effort substitution. A further inspection of the marginal effects for the significant interactions provide some additional insights (results not reported in table). For our interactions, the effect of percent Black students on acceleration gaming turns from being significantly positive (if charter = 0) to negative (if charter = 1) and from significantly positive (if mission specificity = 0) to being zero (if mission specificity = 1).

Some of our control variables yield interesting findings. Generally, effort substitution practices seem to be more prominent among high-performing schools, possibly because these schools have more to lose than those at the bottom. Effort substitution is less likely in larger schools and those in which performance data have

become a part of regular management practices. Larger schools may have a more transparent accountability structure, and schools which use performance data for the purpose of improvement may be less likely to game such data.

Table 6 shows the regression results for our second effort substitution variable – graduation gaming. This time, the rewarded performance measure (graduation rate) is conceptually more closely tied to the unrewarded measures (dropout, attendance, mobility rates). That is, it is more difficult to substitute effort because graduation gains may be linked to gains in the other areas. The table is structured the same way as the previous one. Now, the findings appear to be much more mixed. Only two out of four task demands coefficients show the expected effects. While mission orientation has no independent impact on graduation gaming, it significantly mitigates the reinforcing effect of task demands on effort substitution in two out of four tests. The control variables show no impact, and all models in table 4 perform much worse than those in table 3 regarding their explanatory power.

We further explore our findings by examining whether the effects of our main variables change in different contexts, that is, for different values of our control variables. To do so, we return to our findings from table 3, which shows the expected significant effects on acceleration gaming (a case where rewarded and unrewarded performance measures do not conceptually overlap). We rerun our models and now interact each main variable at a time (percent low-income students, percent Black students, schools with charters, and mission specificity) with all context variables (performance grade, size, climate, focus on performance data, and lack of resources), resulting in four separate

specifications each of which includes one main variable and five interactions with that variable (results not shown).

The main take away is that we do not see much change in coefficients for different levels of the context variables, with a few exceptions. A lack of resources reinforces the effect of task demands on effort substitution, particularly for schools serving low-income families. Further, larger schools seem to be better able to deal with high task demands, as they show a significantly weaker link between task demands (percent Black students) and effort substitution. At the same time, larger schools appear to be less capable of leveraging the benefits of a strong mission orientation. For these schools, mission specificity is significantly less impactful in reducing effort substitution.

3.7 Discussion

In this study, we have examined performance scores of high schools in one school district. By studying one district, we are able to hold all performance-system related factors constant and leverage the organization-level variation among schools. We construct two measures of effort substitution (acceleration gaming and graduation gaming) by comparing whether schools do better on rewarded indicators (career acceleration and graduations rates) as opposed to unrewarded ones (dropout, attendance, and mobility rates). We find consistent evidence that task demands are associated with effort substitution and partial evidence that mission orientation is linked to mitigating it and moderates the task effect. However, we also see significant differences in the results

based on whether rewarded and unrewarded performance measures are conceptually linked or not.

In this section, we discuss our findings and explain their implications and contributions. First, the paper contributes to the development of a tangible theory of performance gaming. While previous research has documented evidence for gaming (e.g., Heinrich & Marschke, 2010; Hood, 2006; Jacob & Levitt, 2003), little is known about organizational factors triggering gaming, particularly explanations beyond the attributes of the performance system itself. Our paper categorizes types of gaming, focuses on one of these types (effort substitution), and develops and tests three specific hypotheses.

Second, our findings help to better contextualize effort substitution theory. Although our two effort substitution measures (acceleration and graduation gaming) are reasonably well correlated ($r=0.48$), our theory about the effects of task demands and mission orientation turns out to work much better for the former, which consists of rewarded and unrewarded performance indicators that are conceptually disconnected. In cases in which such indicators are not much linked, decision-makers are being put in a position to make real choices, where doing more of A automatically means doing less of B. When dealing with trade-offs and tough decisions, managers are more susceptible to external pressure (task demands) and likely to look for a moral compass (purpose). However, when decisions are less controversial, factors like these will be less influential.

Third, the finding that task demands foster effort substitution has important social equity implications. In a government setting, dealing with difficult tasks often means working with underserved populations, which might end up doubly disadvantaged in

performance systems where effort substitution occurs. The first disadvantage is that these groups are usually facing other societal problems outside the control of government agencies. In the context of schools, this might include the effects of crime, domestic issues in the home, and limited access to resources such as transportation and technology, which can all affect student achievement and make meeting performance standards difficult. The second disadvantage comes from what we found in this study – schools with a higher percentage of disadvantaged students are associated with more effort substitution, indicating that such students might be neglected in receiving educational services in unrewarded areas at the expense of the school trying to meet rewarded performance standards.

Our findings for the mitigating and moderating impact of mission orientation were not as consistent as those for task demands. One explanation could be that positive and negative effects of mission orientation can cancel each other out. As expected, in some cases a strong mission orientation may create resistance to gaming behaviors and effort substitution because the latter rarely benefit clients and help accomplish the mission. However, in other cases in which performance metrics are considered to be at odds with the organization’s mission, shortcuts “around the system” could be perceived to be a legitimate strategy to focus on mission as well as meet the formal requirements set by the performance system. Our findings call for further research with respect to reinforcing and conflicting effects of mission versus performance orientations.

Like other studies, our paper is prone to limitations. The first is that our measures of mission orientation are proxies, which are able to capture the existence and content of school charters and missions, but not their implementation. In addition, mission

statements do not capture actual behavior of mission alignment. Measuring the true extent to which a mission statement or charter affects school administrators' decisions and behavior would require a survey of multiple informants for each site with high response rates as well as detailed qualitative observations. Yet, the use of school types and missions statements to distinguish among different degrees of mission orientation is a proven practice (Boerema, 2006; Desmidt, 2016; Garrow & Gursky, 2013; Henig et al., 2005). The second limitation is related to the generalizability of our findings since we study the performance system of one school district as our case. This strategy allowed us to hold all system-related variables constant, but our results may be prone to picking up on some of the idiosyncrasies of the case. However, the observations we make in our case seem compatible with what others have found in other states and school districts (e.g., Bohte & Meier, 2000; Radin, 2006). Furthermore, the performance system we study seems to share typical features of the "average" system: data are linked to consequences but measures are incomprehensive and performance contracts incomplete, hence, allowing for effort substitution and gaming.

We see three specific recommendations for the management of performance systems and the mitigation of effort substitution. First, key performance indicators need to capture all relevant performance dimensions. While no performance system is "perfect" or able to measure everything (in fact, too many indicators can be counterproductive as well), the selected indicators should be broad enough to capture changes in all mission-relevant, but also important non-mission based (social equity, procedural justice, social trust etc.), areas. Performance systems can still be strategic and prioritize certain objective over others, but if managers are not supposed to generate gains

in rewarded areas at the expense of unrewarded ones, then system designers need to attach at least some rewards to non-priority targets and indicators that measure negative unintended consequences.

Second, benchmarking needs to account for varying task demands. Administrators tend to engage in effort substitution practices if tasks are difficult and performance targets perceived to be unachievable through other managerial means. Hence, absolute targets and across-the-board minimum standards can be counterproductive, and perceptions of stretch targets vary across contexts. Systems could attach higher rewards to relative performance gains (using an “added-value” perspective) than overall level-improvements. Or, organizations could be divided into benchmarking clusters based on the task demands they are facing, rather than employing league tables that compare all organizations in the same ranking.

Third, performance systems need to be accompanied by norms that emphasize the organizational mission and purpose. The understanding of, and identification with, an organization’s purpose has the potential to mitigate against managerial short cuts and dysfunctional behaviors such as effort substitution. While helpful to meet targets in the short-term, these behaviors are likely to harm clients and reduce impact in the long-run. Hence, performance system designers need to ensure that metrics and purpose are not perceived to be disconnected from each other. Rather, systems need to emphasize that performance metrics are supposed to support, visualize, and improve the organization’s value.

3.8 Conclusion

Examples of gaming and cheating as a response to performance systems have become too common to be dismissed as outliers or exceptions. This paper contributes to a theory of performance gaming, as it studies why public organizations engage in effort substitution (i.e., directing effort towards rewarded as opposed to unrewarded areas). We argue that effort substitution becomes more likely if tasks are difficult; less likely in the presence of a strong mission orientation; and that mission orientation is linked to mitigating the task demands effect. We test these hypotheses by examining a five-year panel data set of 64 high schools in one school district, which allows us to hold all performance system-related factors constant.

We find support for the effect of task demands and partial support for the main and interactive effects of mission orientation. We think this is intriguing since dealing with difficult tasks in a government setting often means working with underserved populations, which may end up doubly disadvantaged if improvements in one area are traded for null effects in another. In particular, we see support for the hypotheses when rewarded and unrewarded measures capture different dimensions of performance and, hence, put decision-makers in a position to make real choices. Results are mixed when rewarded and unrewarded measures are conceptually linked. Our findings suggest that our theory holds when stakes are high but that task demands and purpose are less influential driving forces when decisions are less controversial.

IV. ESSAY 3: DOES PROSOCIAL IMPACT REDUCE PERFORMANCE DATA
GAMING? THE ROLE OF DATA VISUALIZATIONS AND EXPERT-NOVICE
DIFFERENCES

4.1 Abstract

While research has documented the framing effects of performance data on decision-making, little is known about the impact of data visualizations on performance gaming. We propose that prosocial impact information (when qualitative and visualizing benefits for clients in need) reduces gaming. We also point to expert-novice differences and suggest that impact information is less influential among experts, with some exceptions. We conduct an experiment with samples of citizens (novices) and school leaders (experts) in which subjects have to decide whether they are willing to omit poorly performing students when calculating school-wide performance scores. We find that school leaders game less than citizens, and that the former are less responsive to visualizations of prosocial impact data cues than the latter. However, impact information becomes more influential if it can grab school leaders' attention and if leaders have little work experience. More broadly, our findings suggest being cautious with generalizations across citizen and manager samples. So, there is evidence to suggest that when performance data is presented with a prosocial perspective, public managers and citizens will engage in less performance gaming. However, public managers and citizens differ in how they perceive performance data and under what conditions.

4.2 Introduction

Research on performance information use has documented how design choices, framing effects, and cognitive biases affect public managers' decisions and behaviors (Moynihan et al., 2017; Andersen & Moynihan, 2016; Webeck & Nicholson-Crotty, 2019). While findings explain, for example, how different ways to present performance information alter managerial decision outcomes, little is known about dysfunctional data use: to what extent do differential visualization of the same information mitigate performance gaming or cheating?

Meanwhile, the literature on performance gaming is growing, suggesting that most incentive systems also yield unintended, dysfunctional responses (Eterno & Silverman, 2012; Soss et al., 2011), which can be highly detrimental even if such responses are not dominating the entire system (Jacob & Levitt, 2003; Kelman & Friedman, 2009). To that effect, scholarship in public administration has examined how organizational factors and system configurations influence gaming and cheating (Benaine & Kroll, 2019; Heinrich & Marschke, 2010), with lesser attention devoted to cognitive and value-driven mechanisms.

In this paper, we aim to bridge these two literatures. We examine how the different framing of performance information regarding its prosocial impact affects the odds of public managers, who are exposed to this information, to engage in gaming behaviors. Based on previous work in this area, we propose a general model, where prosocial impact information, specifically when qualitative (as opposed to quantitative) and when visualizing benefits for clients in need, reduces performance gaming. Our main

rationale is that seeing vivid visualizations of one's prosocial impact makes people engage in behaviors that can maximize effect, rather than undermining potentially real impact through data manipulation. At the same time, we argue that general framing effects of information may matter more for novices as opposed to area experts whose behavior is a function of specialized knowledge and professional socialization. We suggest two adjustments to the general model. Prosocial impact information is likely to reduce performance gaming among experts if experts are still relatively inexperienced and if the data can grab experts' attention.

To test our theory, we conduct a randomized experiment with samples of citizens (novices) and school leaders (area experts). We put subjects in a hypothetical scenario in which they have to decide whether they want to omit individual students when calculating a school-wide performance score. Subjects will be considered "successful" in our scenario when they meet a predefined target, whose achievement essentially requires the omission of the weakest students. At the same time, subjects are notified that omitted students will be transferred out of the school, taking away their chance to participate in the school's student improvement program. All subjects are randomly divided in four groups, receiving different visualizations of positive performance information (PI) for the student improvement program (quantitative PI, qualitative PI, qualitative PI emphasizing impact on Black students, and no PI). We create a measure of performance gaming as our dependent variable by adding the number of students omitted, weighted by how poorly a student was performing.

In line with our model, we find significant differences between ordinary citizens and school leaders. While qualitative impact information, particularly when visualizing a

positive effect for Black students, makes it less likely for citizens to engage in performance gaming, we see no such effect for the school leader group. When further examining the group of school leaders, we see that impact information is more influential the lesser experienced the school leader is. Further, we find that the more attention that group devoted to studying the impact information, the larger its mitigating effect on performance gaming.

4.3 General Model of How Prosocial Impact Reduces Performance Gaming

In this section, we propose a general theory of how prosocial impact information is linked to performance gaming. In doing so, we draw on previous studies that have examined the framing effects of performance data and prosocial impact conditions (e.g., Belle, 2013a; Grant, 2008a; Olsen, 2017), although none of these studies utilized performance gaming as the outcome variable. In the subsequent section, we suggest specific adjustments to our general model that account for the differences between novices and experts as potential data users.

We understand performance gaming as generating positive performance data without achieving the indicator's underlying objective. Gaming can be done through manipulating data or resources (Benaine & Kroll, 2019). One specific type of gaming we turn to in this study is cream skimming. This involves selecting clients that are easier to serve as well as underserving undesirable clients or excluding them from performance metrics or reports (Bohte & Meier, 2000; Koning & Heinrich, 2013). We define prosocial

impact information as performance data that link public service actions and outputs to measurable impact on clients or citizens. Like indicators for short-term outcomes, prosocial impact information tends to portray the immediate effect on clients, and it captures impact descriptively rather than being able to prove causality. Unlike short-term outcome data, we expect prosocial impact information to measure substantive changes in behaviors or achievements, excluding indicators for lower-order outcomes such as program participation or satisfaction. In that sense, all prosocial impact data can be considered outcomes, but not all outcomes are necessarily impact data (Hatry, 2006; Van Dooren et al., 2015). In this paper, however, we use the two terms, prosocial impact data and performance information, interchangeably to be able to connect to previous research that has mostly employed the latter term.

A great deal of the literature on performance systems borrows its assumptions from agency theory, suggesting that better specified – more complete – contracts with measurable outcomes may constrain opportunistic behavior among agents (Kettl, 1997; Heinrich & Marschke, 2010). Research on prosociality offers an alternative mechanism to align the individuals' interest with the organizational mission (Bolino & Grant, 2016). If bureaucrats are aware of the prosocial impact they have, they will likely put more effort into their public service work. Along these lines, perceived prosocial impact can motivate purposeful in-role and extra role behaviors as well as mission-based performance. Consider the following examples.

High levels of prosocial impact make public managers care about, and act upon, performance data (Moynihan et al., 2012). Increased prosocial impact created through interactions between public service providers and direct beneficiaries increased the

providers' long-term productivity significantly (Grant, 2008a). Similarly, Belle (2013a) observed prosocial impact in nurses' personal encounters. When nurses, who made healthcare kits, saw healthcare workers using these kits, the nurses made fewer errors and demonstrated higher productivity in their work. Cumulatively then, prosocial impact seems to motivate bureaucrats to increase performance through real improvements, rather than increasing performance scores on paper via the manipulation of data or effort.

Acting prosocially and ethically when conducting government work are behaviors, which are both based on the same principle: fostering the public interest (Wright et al., 2016). Accordingly, research shows that unethical behavior is more related to self-interest instead of people's other orientation (Gino, 2015), and prosocial motivation specifically has been found to be at odds with lying in a set of dice game experiments (Olsen et al., 2018). Further, in line with social exchange theory, there is evidence that seeing ethical behavior and morality in others triggers one's own prosocial behavior (O'Keefe et al., 2018; Zhang et al., 2017).

Overall, we argue that whether managers respond purposefully or dysfunctionally to targets and requirements is a function of how meaningful they perceive performance information to be. If data can portray the immediate impact of one's work on others' lives, managers may consider such data as instructive and helpful rather than a means with no end or a part of a technocratic compliance exercise. Put another way, prosocial impact information provides meaning and makes performance systems less likely to appear as an external burden with little utility, which could otherwise serve as a justification to manipulate data and performance.

H₁: Prosocial impact information reduces performance gaming.

Impact information, like other performance data, can be presented and portrayed in several ways. For the most part, public agencies report performance data in a quantitative and aggregated format following a predefined reporting routine, although we know that a great deal of performance feedback that agencies act upon is nonroutine qualitative information (Kroll, 2013). Along the same lines, government performance information has been differentiated in statistical (“hard numbers”) and episodic (“stories about specific cases”) information (Olsen, 2017).

In his classic critique of the utility of management information systems, Mintzberg (1975) argues that performance information created through such systems is mostly simplistic and backward-looking and, hence, inferior to soft and rich qualitative feedback that can be collected on the spot. Similarly, research in public administration has shown that public managers prefer nonroutine, qualitative feedback as opposed to routine, quantitative data (Kroll, 2013), although the use of both information types can complement each other (Tantardini, 2019). Using a set of experiments, Olsen (2017) shows that subjects are better able to recall episodic, qualitative information about public service performance compared to quantitative data because the former invokes greater emotional responses. In our conceptualization of qualitative impact information, we certainly include images, such as those of public service beneficiaries, acknowledging that images and symbols have proven to show a significant effect on how the performance of public service provision is perceived (Alon-Barkat & Gilad, 2017). Overall, we argue that qualitative impact information is able to provide a narrative for

specific cases, which is often more convincing and relatable than averaged or aggregated large-scale data.

H₂: Qualitative prosocial impact information is more influential than quantitative impact data in reducing performance gaming.

A second qualification of the effect of prosocial impact information on performance gaming is related to the characteristics of the client group at which the positive impact is directed. Specifically, we propose the effect of impact information increases if the beneficiaries of the public service will be considered historically disadvantaged. We think this is the case because public values, such as racial equity and minority protections, which are also known to shape the behavior of bureaucrats (Bozeman & Johnson, 2015) are much more widely supported today than decades ago (although the ways how to achieve such objectives are still discussed controversially, see Banerjee, Gupta, and Villeval 2018; Dynarski 2018). To that effect, race is described as being “more persistent and more influential” than any other demographic in public policy (Meier, 2016, p. 9).

Research shows that Americans generally support helping the poor but disagree on what government’s role should be and how help should be given (Howard et al., 2017). This holds true despite the fact that motivations to help as well as general attitudes towards disadvantaged groups can be shaped by stigma, racial paternalism, and varying perceptions of deservingness as well as double standards (Maynard-Moody & Musheno, 2012; Oorschot, 2000). For example, children are often considered more deserving than adult groups of historically disadvantaged populations (Meanwell & Swando, 2013; Will,

1993). Most broadly, we propose that prosocial impact information is particularly influential if the client group they report on is considered disadvantaged.

H₃: Qualitative prosocial impact information becomes more influential if it portrays benefits for historically disadvantaged groups.

4.4 Theoretical Adjustments for Expert-Novice Differences

An important point in our argument is that we propose adjustments to our model based on expert-novice differences. While so far mostly neglected in the literature, we suggest that the visualizations of prosocial impact data cues outlined above are less influential if decision-makers are area experts as opposed to novice decision-makers. We argue that expert decision-making is a function of expertise, professional norms, and socialization, and it is less susceptible to different presentations of the same impact information. Put differently, we expect prosocial impact information to be generally less influential in mitigating performance gaming behavior for managers who are area experts.

Experts' behavior is often driven by professional norms and socialization (Grøn, et al., 2019; Moyson et al., 2018). Expertise feeds into naturalistic decision making, where experience creates a knowledge base for shaping decisions (Rosen et al., 2009). Recent research shows that decision-making can be based on intuition and experience over analytical sources (Hine et al., 2018). Experts may develop greater dependence on their own expertise because they have also developed greater knowledge-sharing

mechanisms through their social interactions and have become more accustomed to the norms of their organization (Manaf et al., 2018).

With regard to performance information use, there is scattered evidence showing that information preferences among area experts (administrators, internal stakeholders) and the general public (citizens, external stakeholders) differ, in that the former prefer feedback about processes and effectiveness, while the latter favor data on equity and information drawn from opinion surveys (Walker et al., 2018; Woolum, 2011). If experience matters, it does in a way in which managers with less experience are more likely to consider and act upon performance data (Kroll, 2015). Experts might think that they know already everything they need to know, constraining their information screening behavior and making it prone to confirmation bias and patterns of motivation reasoning (Kahan, 2016). While often being less data-driven than novices (Hershey & Walsh, 2000), experts are known to make effective intuitive decisions based on specific knowledge, pattern recognition, and automaticity (Salas et al., 2010). Overall, we argue that decision-making tendencies among experts and novices vary, and that we expect prosocial impact information to be less effective in shaping experts' behavior.

H4: The negative effect of prosocial impact information on performance gaming is stronger for novices than experts.

A second theoretical adjustment is that prosocial impact information is more influential if it can grab experts' attention. Business research has argued that information flow and the direction of attention are as important to understanding firm behavior and performance as factors such as resources or capabilities (Barnett, 2008; Ocasio, 1997).

Once experts pay attention to impact information, they are likely to get more out of it than novices. They are better able to make sense of such feedback because they possess the knowledge base to contextualize new pieces of information, and they are able to make quick but accurate judgements (Salas et al., 2010). Along these lines, when devoting attention to impact information, “more experienced managers can be expected to place more weight on relevant (functional) cues and less on irrelevant (peripheral) cues” (Perkins & Rao, 1990, p. 2). Expert managers can get a great deal out of impact information and scorecards (Naranjo-Gil, 2009), but our point is that such an effect only materializes if experts actually consider impact information when making decisions. When experts pay more attention to information, it will play a larger role in their decision-making than it would have been if the experts were novices. The literature shows that experts are better at sense-making of information (Salas et al., 2010) and so we argue that experts’ attention will be important in reducing performance gaming when added with prosocial impact.

H5: The negative effect of prosocial impact information on performance gaming is reinforced, the more attention expert users devote to the information.

4.5 Data and Methods

We have set up our study of gaming in the area of education, and specifically schools, for two reasons. First, the school system is fairly representative of other multi-level agency systems within the public sector, and it has experienced a great deal of

performance measurement and gaming (Jacob & Levitt, 2003; O'Toole & Meier, 2011). Second, due to real-life exposure, most individuals in a general subject pool can relate to challenges and decision scenarios that draw on concepts such as schools, testing, and students. While we will describe our samples in more detail further below, we want to acknowledge upfront that we work with two samples: citizens, who we consider novices when it comes to decision-making on the school system, and school leaders (professionals) who serve as our proxy group of area experts. In our appendix, we show screenshots from the online experiment. We have not attached the entire survey but share what we think are the most relevant screens.

4.5.1 Capturing Performance Gaming

To measure gaming, we made subjects work through a fictive scenario at the end of which they had to make choices. Participants of our survey experiment play the role of a school administrator who has the task of deciding whether to keep students or transfer students to another school. The fictional students are academically low-performing students whose scores have brought down the overall school average and have created a situation, where the school has not met the district target of a 70% overall school average.

The incentive to transfer these students to another school is that the school's overall average will increase, allowing the school to meet the district target. The disincentive to transferring these academically-challenged students is that sending them to another school will prevent them from accessing a successful reading program, SuperStudent, which is only available at the current school (SuperStudent is a fictional program, created for purposes of this experiment, but such programs exist in US schools).

The dilemma of deciding whether to keep students or transfer them (and, hence, help the school administration or the struggling students) presents the participants in this experiment with an opportunity for gaming the system: school-average metrics can be improved at the expense of the individual. Appendix 1.A shows two screens we used to set up the scenario.

At the end of the scenario, participants were provided with a list of students and their test scores, and they had to decide which students to keep or transfer (see appendix 1.C.1). Based on participants' decisions we calculated a gaming score, which tends to be high if students with low test scores were omitted. Our gaming variable ranges from 0 (no gaming) to 16 (meaning complete gaming). The participants receive gaming scores depending on the student that is transferred. For example, a student with a very low standardized test score of 17% provides participants with a gaming score of 3, whereas a much higher-performing student (a score of 63%) provides participants with a score of 1 if transferred (see appendix 1.C.2). Gaming points are only given to participants if the participant transfers a student, whereas keeping a student is scored 0. Keeping a student is not considered gaming because the actual performance of the overall school is not being distorted. This measurement of gaming behavior is similar to the one used by Bohte and Meier (2000) who measure gaming as the percentage of students that were exempted from taking standardized tests.

4.5.2 Treatments and Other Variables

As described above, participants need to decide whether they want to keep or transfer low-performing students. Keeping students, in our fictive scenario, also means being able to help struggling students via enrollment in a reading program called SuperStudent. To examine whether, and to what extent, prosocial impact information affects performance gaming, different groups of participants were subjected to different visualizations of performance information about the success of SuperStudent before being asked to make any decisions. The experiment employs three treatment groups and one control group (the latter did not receive any performance information about SuperStudent). All three treatment groups received positive feedback about the impact of the program on students. However, the “quantitative PI” group received only quantitative data: two tables that show performance scores for the SuperStudent school over four years and compared to another school. The other two treatment groups received qualitative information in the shape of testimonials as well as pictures of individuals who were positively impacted by the program. Here, groups were differentiated based on the race of the impacted students and stakeholders. That is, while the text was identical, pictures and names were traditionally African American for one group (“qualitative PI [black]”) and Caucasian for the other (“qualitative PI [white]”). All exact treatments are featured in appendix 1.B.

Participants were randomly assigned to one of the four groups. We performed a manipulation check for our treatments. We used three items that pick up on the perceived social impact of the treatment (one sample item is: “The information I have reviewed indicates that SuperStudent is benefiting students”). Correlational analysis shows that our

three-item measure of perceived impact is – across samples of citizens and professionals – significantly positively correlated with the PI treatment groups and negatively with the (no PI) control group. In line with recent recommendations by Mutz, Pemantle and Pham (2019), we – instead of presenting mostly uninformative balance tests – estimate treatment effects with and without a set of control variables to our experimental models. We do provide a balance table in Appendix 3 to show that the different groups were balanced and one group did not have significantly more of one variable than another group.

One additional variable with relevance for our hypothesis testing is school leaders' inexperience, which is captured via a reversed ordinal 5-point scale of work experience. We use this measure to differentiate among professionals with more or less work experience. A second variable that is of theoretical interest is the attention participants devoted to the performance information they received. This variable is measured as the time (in seconds) that subjects spent looking at the treatment screen. Additional control variables include minority (everyone who identifies as non-white and non-Hispanic); female (coded zero for males); age (using brackets); and liberal political ideology (a scale where zero is labelled “conservative” and ten is labelled “liberal”).

4.5.3 Samples of Citizens and Professionals

To capture groups of novices and experts, we use samples of citizens and professionals (school leaders) for our survey experiment, which is set up in the area of test scores and schools. The experts (school leaders) are experts in the field of education and should be more familiar with the roles being asked of them compared to the novice

(general population sample). We can expect the experts to be experts because they professionally work in their field and we expect the general population to have less expertise as a collective because they were not randomly pulled from a professional pool. Our sample of school leaders stands at 286 (this number is lower for some multivariate analyses due to missing values on some of the survey questions). This group includes principals, assistant principals, lead teachers, school counselors, program directors, and district administrators, spanning elementary school through high school, and spanning charter and traditional public schools. All school personnel are from the same public school district.

To collect this data, we first received approval from our institutional review board and the school board's research approval committee. Being that our experiment was completely online, we contacted school leaders via e-mail. We accomplished this by dividing the district into its major sub-areas. We then visited schools or scoured school web sites for publicly available e-mails. We eventually contacted every school whose information was available, reaching a total of about 4,200 school leaders. We sent about reminders during our survey period ranging between September 2018 and April 2019. Like most other randomized controlled trials, our focus is with internal rather than external validity. That is, it was important to us to recruit real-life school leaders rather than being able to argue that this group of volunteers is representative of a specific school district. To ensure that the school leaders, who otherwise had no personal stake in the fictive decision scenario, feel more invested in the experiment, we offered the following incentives: Meeting the fictive 70% target of the case scenario resulted in the receipt of bookstore gift card for a 20% discount, while retaining students and giving them the

opportunity to improve triggered a book donation to schools in need if a majority of school leaders chose the latter option. (Eventually, we donated books to three schools.)

Our citizens sample stands at 758 participants. In December 2017, we recruited subjects via the Amazon Mechanical Turk (M-Turk) platform that has been found to be largely representative of population-based random samples (Coppock, 2019; Huff & Tingley, 2015). It has also been used in public management research that focuses on citizen attitudes and behaviors (for an overview, see Stritch et al., 2017). We paid \$0.50 to each participant as well as a fee to Amazon. Participants were rewarded for participating and not based on how they answered questions in the experiment.

To make sure that participants paid attention when filling out the survey, we used the following two strategies. First, we constructed the survey in a way, where subjects could only complete the sorting task by dragging and dropping each individual student into either the “keep” and “transfer” categories. Hence, not doing anything on the decision screen was not a viable option if one’s goal was to complete the survey and get paid. Second, we reached out to a random sample of participants that completed the survey with above-average speed and asked for more explanations of their decision-making.

We wanted most of the sample to be from North America in order to be somewhat comparable to our school leader group. At the same time, we do not think that our model is specific to the US or North America, which is why we find it beneficial to test our theoretical expectations using a more global sample. Our final sample is 80% North American and 20% international (mostly from Asia/India). Since our sample suggests

that there was more performance gaming outside of North America, we control for this variable in our models. Further, we interacted a North America dummy variable with the main effects presented in this paper but find no significant differences (results not shown). Hence, our findings for the citizen sample hold beyond North America. Table 7 shows the descriptive statistics of both our citizen sample and our professionals sample.

Table 7: Descriptive Statistics

Variable	Citizens Sample			Professionals Sample		
	Mean	Std. Dev.	Range	Mean	Std. Dev.	Range
Gaming Score	6.72	4.61	0 – 16	4.38	4.76	0 – 16
Quantitative PI	0.25	–	0 – 1	0.27	–	0 – 1
Qualitative PI (students generally)	0.26	–	0 – 1	0.21	–	0 – 1
Qualitative PI (Black students)	0.24	–	0 – 1	0.28	–	0 – 1
Attention	65.26	56.50	1.34 – 640.96	373.14	4,211.09	11.10 – 71,239.96
Minority	0.41	–	0 – 1	0.78	–	0 – 1
Female	0.49	–	0 – 1	0.78	–	0 – 1
Age	2.39	1.09	1 – 5	2.96	0.81	1 – 4
Liberal	6.06	2.77	0 – 10	5.67	2.44	0 – 10
North America	0.80	–	0 – 1	–	–	–
Inexperience	–	–	–	2.10	1.27	1 – 5

The descriptive statistics for both our samples can be found in table 1. On average, professionals spent about two minutes⁵ looking at the treatment, while citizens did so for about one minute. Within the citizen sample, 41% of the individuals identified as racial minorities as opposed to 78% in the professional sample. Roughly half of the citizens are female, while this number among school leaders is 78%. The professionals (46% are between 41 and 54) are on average older than the citizens (47% are between 26 and 35). Both samples seem to be fairly centrist regarding participants' political views – citizens scored 6.1, professionals 5.7 on an ideology scale, where ten was labelled liberal and zero was labelled conservative. While we did not expect citizens to have any work experience in the school system, 45% of professionals reported 21+ years of experience, followed by 28% reporting 15-20 years of experience.

4.6 Results

Table 8 provides a test of hypotheses 1-3 about different visualizations of impact data as well as differences between novices (citizens) and experts (school leaders) (hypothesis 4). The table shows the effects (OLS regression coefficients) for the three treatment groups (the control group serves as the reference) on individuals' gaming scores. The first two models combine the two qualitative PI groups into one, while models three and four separate the qualitative treatment by target group (white versus black). The table contrasts findings across samples (citizens versus professionals), and it

⁵ Table 1 shows the larger, unadjusted mean that, however, is driven by two extreme values.

accounts for a set of demographic controls, although the results remain the same with or without control variables (results without controls are shown in appendix 2.B).

Table 8: The effect of prosocial impact treatments on performance gaming

	Citizens (1)	Professionals (2)	Citizens (3)	Professionals (4)	
Quantitative PI	-0.43 (0.34)	-0.66 (0.41)	-0.43 (0.34)	-0.66 (0.41)	
Qualitative PI (combined)	-0.93 (0.02)	-0.03 (0.96)			
Qualitative PI (white students)			-0.68 (0.14)	-0.00 (0.99)	
Qualitative PI (black students)			-1.20 (0.01)	-0.06 (0.94)	
Constant	10.85 (0.00)	5.83 (0.00)	10.85 (0.00)	5.84 (0.00)	
	N	758	263	758	263
	R ²	0.088	0.015	0.089	0.015

Note: p-values in parentheses (significant results marked in bold for hypothesized effects); control group serves as reference category; control variables: Minority, Female, Age, Liberal, North America (last one only in citizens sample); robust standard errors.

Hypothesis 1 suggests that prosocial impact information reduces performance gaming. We largely see evidence for this claim, as the gaming scores for all treatment groups are lower than that of the control group (although not all coefficients are statistically significant). Regarding types of data visualization, we find that qualitative impact information is more influential in reducing gaming than quantitative data (model 1), which supports hypothesis 2. Further, we see in model 3 that when broken down by groups of beneficiaries, impact information reduces gaming behavior to a larger extent if

program beneficiaries tend to belong to a historically disadvantaged group (black versus white students). This supports hypothesis 3. At the same time, we note a stark contrast across our two samples. The findings that we just described only hold for the citizen but not the school leader sample (models 2 and 4). These results are largely in line with hypothesis 4, which emphasizes that the general model outlined in our first three hypotheses may work for novices but not area experts. In fact, we proposed adjustments to such a general theory when it comes to professionals, and we examine these adjustments next.

One notable difference between novices and experts can be found in table 7. On our scale ranging between 0 and 16, school leaders (4.4) gamed significantly less than citizens (6.7). In essence, 47% of all school leaders and 28% of all citizens did not game at all.

In addition to the differences documented in Tables 7 and 8, Table 9 provides further insights into some idiosyncratic effects for the expert group. This table draws only on the sample of school leaders, and it interacts the treatment effects with our inexperience variable (the latter being a reversed measure of work experience). Using this variable allows us to perform a second test of our novices-expert distinction hypothesis. We now differentiate among more and fewer experienced experts, assuming that, if our hypothesis holds true, experts with less experience will be more influenced by impact information. The modelling here is otherwise similar to our strategy in Table 8. It is noteworthy that the attention and inexperience variables were mean-centered, so that the coefficients in the first five rows show the treatment effects when attention and

inexperience are at their respective means. Our focus, however, is on the interaction terms.

Table 9: Moderations of the effect of prosocial impact (school leaders)

	(1)	(2)
Quantitative PI	-3.92 (0.01)	-0.81 (0.33)
Qualitative PI (white students)	-3.93 (0.09)	-0.16 (0.85)
Qualitative PI (black students)	-2.60 (0.04)	-0.33 (0.69)
Attention	0.01 (0.06)	
Inexperience		0.66 (0.23)
Interactions		
Quantitative x Attention	-0.01 (0.01)	
Qualitative (white) x Attention	-0.01 (0.08)	
Qualitative (black) x Attention	-0.01 (0.01)	
Quantitative x Inexperience		-1.08 (0.10)
Qualitative (white) x Inexperience		-1.36 (0.04)
Qualitative (black) x Inexperience		-0.20 (0.76)
Constant	7.79 (0.00)	5.92 (0.00)
N	263	263
R ²	0.040	0.037

Note: p-values in parentheses (significant results marked in bold for hypothesized effects); control group serves as reference category; control variables: Minority, Female, Age, Liberal; robust standard errors. The Attention and Inexperience variables were mean-centered to allow more straightforward interpretations of the main effects.

Model 2 in Table 9 shows that, generally, impact information reduces gaming more if experts tend to be inexperienced. While all three coefficients for the interaction

effect show the expected negative sign, only one is statistically significant at $p < 0.05$ and another is just outside of the $p < 0.10$ cut-off. What this suggests is that by and large the same pattern we find across novices and experts also holds when differentiating among more and less experienced experts. This is more evidence for hypothesis 4.

Model 1 in Table 9 tests an extension of our theory captured by hypothesis 5. While experts may be less impacted by performance information, they are able to get much out of such feedback if they choose to pay attention to it. As attention increases, we see that the mitigating effect of all types of data visualizations on gaming behavior becomes stronger (more negative). Two coefficients are significant at $p < 0.05$, whereas one is at $p < 0.10$. While the size of the coefficients appears to be small at first glance, we need to keep in mind that attention was operationalized as the time the treatment screen remained open in seconds. That is, for each additional minute devoted to studying impact information, the gaming behavior will be reduced by 0.60. This finding supports hypothesis 5.⁶

4.7 Discussion

Visualizations of prosocial impact data, particularly when qualitative and visualizing benefits for clients in need, can lead to less performance gaming. We find that this theory holds for our general population sample. At the same time, these findings differ for our area experts, a group whose behavior was largely unaffected by different

⁶ As a check, we also ran the same tests for the citizens sample, but here more exposure to information does not yield a significant interaction effect (results shown in Appendix 2C).

visualizations of prosocial impact data. When dividing that group of school leaders by work experience, the same general pattern emerged: Experienced leaders are significantly less likely to be swayed by performance data than less experienced leaders. More descriptively, school leaders also engaged much less in gaming behavior than citizens. Overall, we documented stark contrasts between novices and experts.

We explain novice-expert differences due to socialization on the one hand, and a differential knowledge base on the other (Manaf et al., 2018; Moyson et al., 2018; Salas et al., 2010). Expert opinions may be much firmer than those of novices. Empirically-grounded experience – and values developed through experience – have shaped experts' mindsets over a long period of time, which is why we are not surprised that school leaders gamed much less than our citizen group. The former is aware that gaming the system, while widely practiced, is not normatively desired. The fact there was less gaming among experts is likely attributable to an organizational culture effect, which we could not account for since all expert participants came from the same school district. It is possible the participants from the school district sample are familiar with the scenario and are less susceptible to gaming because of experience with how to deal with such scenarios. Additionally, experts have often been exposed to a host of relevant information that can be transferred across similar decision situations. Experts may be more prone than novices to give a socially desirable response when it comes to decision dilemmas, possibly shaped by the norms of their profession. Also, the behavior of experts may be shaped by their professional socialization and what they deem acceptable as a collective. This type of socialization base might not be as developed in novices. Hence, the marginal

impact of one more piece of information (such as in our feedback treatments) is not likely to be pivotal in changing decision outcomes or routine behaviors.

One would expect that public values such as diversity and social equity, which are largely featured among the current generation of public managers (Hill & Lynn, 2015), affect professionals' behaviors – instead, school leaders did not alter their decision when confronted with performance information showing potential improvements for African American stakeholders. Specifically, since most of the school leaders in our sample belong to a minority group themselves, we would suspect to see a more responsiveness towards minority students (Meier, 2019). One explanation certainly is the fact that a fair number of leaders in our sample are white Hispanics as opposed to African Americans. But even when interacting the 'qualitative PI treatment targeted towards black students' with black school leaders, we do not see any significant effects (results shown in Appendix 2D).

Our findings also show that experts can get a great deal out of performance data if they choose to pay attention to a new information tidbit. This finding is not limited to qualitative feedback, as it was for the citizen sample, but quantitative data as well. Experts are certainly equipped to make sense of such information, but we will need more research into the factors that make them pay attention in the first place. A second avenue for future research is building a body of literature on the effects of data framing and visualization on the dysfunctional uses of performance information, which tend to be mostly neglected in experimental public management research. To follow the route of scholarship on purposeful data use (e.g., Moynihan et al. 2017; Webeck & Nicholson-Crotty, 2019), work on performance gaming could benefit from adopting some of the

experimental methods employed in research on unethical behavior (Bellé & Cantarelli, 2017a; Hilbig & Hessler 2013).

We recognize that conducting a survey experiment certainly constitutes a limitation. A major concern with such experiments is that text vignettes may not be powerful enough to invoke behavioral or even attitudinal effects. We addressed the issue by using a full case scenario that subjects had to work through before making a decision, rather than simply exposure to one vignette description. We also found significant differences across treatment and control groups, suggesting that our scenarios were successful in triggering varying responses. While we agree that laboratory experiments have certain advantages over survey experiments, we also want to acknowledge that it becomes almost impossible to recruit a sample of 200+ school leaders to participate in a lab study, often leaving researchers with student samples in such settings. We also acknowledge the limitations of using the citizen sample as a proxy measure for novices. Actual novices would be assistant teachers, but the difficulty in gathering a large sample of assistant teachers is similar to that of school leaders. The citizen sample then served as an appropriate proxy measure.

4.8 Conclusion

Using samples of citizens and school leaders, the paper documents the effects of prosocial impact information on performance gaming, and it explores important expert-novice differences. In line with a general theory, we find that impact information reduces gaming among novices if information is qualitative in nature and particularly visualizes benefits for clients in need. However, these findings cannot be confirmed for our group of

experts, suggesting that this group is less prone to differently manipulated information cues. This pattern holds when we compare experts with more or less work experience. Those with more experience are less susceptible to impact information, no matter the type of visualization. At the same time, impact information is influential if it can grab experts' attention, and this includes qualitative as well as quantitative feedback.

A first contribution that our paper makes is to a broader theory of performance information use. While the number of studies on responses from citizens and decision-makers is still increasing (for an overview, see James and Olsen 2017; Moynihan et al. 2017), only very few studies tested the same model on different populations simultaneously. We think that such testing is paramount for broader theory-building because it helps distinguishing the elements of our models that are transferrable across, or unique to, populations such as managers, politicians, citizens, clients, and other stakeholders. Specifically, our findings suggest being careful with generalizations across samples. Effects largely varied between citizens and professionals (school leaders), indicating that extrapolating from one group to the other may be a slippery slope when it comes to responses to performance data. While we find it plausible that certain effects and responses may mirror each other across populations, we propose that taken-for-granted assumptions about their transferability ought to be tested.

For practice, the main conclusion we draw is that performance information needs to be tailored to the needs of different user groups. For novices, information was most influential when qualitative and when impact on specific client groups was emphasized. With experts, the challenge is to direct this group's attention to a new information; and strategies to accomplish this certainly require additional research. However, once experts

pay attention to information, both qualitative and quantitative visualizations are similarly impactful.

Another theoretical contribution is in the area of prosociality and unethical behavior. Interestingly, prior work documented mixed findings where, in some cases, prosociality led to more and in others to less unethical behavior (Bellé & Cantarelli, 2017b; Bolino & Grant, 2016; Olsen et al., 2018). Previous work has argued that decision-makers exposed to prosocial impact information may engage in unethical behaviors if such behaviors benefit clients. Our study provides a reversed test of this argument. We told subjects that gaming (meeting the school target) will harm clients (individual students): So, in a setting in which it is clear that unethical behavior has direct negative consequences for others, prosocial treatments will yield less unethical behavior.

A final contribution is to the literature on performance gaming. A great deal of gaming research in public administration draws on principal-agent assumptions and considers self-interest as the main motive behind actions in the workplace. Hence, research is concerned with accountability systems, incentive structures, and system configurations. Our take is different, in that we start from a prosocial perspective: if we are aware of the direct impact of our work on others, then maximizing this impact can become a more instructive motive than maximizing personal gains at the expense of others. While different visualizations of impact data may also call for different system configurations, the centerpiece of this perspective is leveraging – and strengthening – individuals' altruism, empathy, and public values.

V. CONCLUSION

In this section, I will bring together my three essays for a synthesized cross-analysis. In the literature essay that introduces my dissertation, I provide an explanation for other avenues for further research, specifically in the realm of experimental studies, where we can establish stronger causal links between gaming behavior and hypothesized factors. I activate some of my proposed ideas that I presented in the literature review, by examining other-oriented constructs, such as mission-orientation (in the observational study) and prosocial impact information (in the experimental study). Both empirical studies and the literature review examine performance gaming as the variable of interest and the influence of other-oriented constructs as antecedents that reduce performance gaming. The theoretical contribution is the importance of other-oriented constructs as antecedents in reducing performance gaming. For the practical implications for public management, other-oriented constructs can provide mechanisms for aiding public managers in reducing gaming and improve decision-making.

5.1 Summary of Conclusion

In public organizations, performance systems underscore the importance of democratic accountability and government responsiveness to the public interest (Van Dooren & Van de Walle, 2008). But performance systems bring with them dysfunctional behavior, such as performance gaming, where individuals and organizations seek to undermine organizational goals by manipulating performance output in perverse ways. Principal-agent theory is the traditional explanation for performance information use (Jensen & Meckling, 1976). This theory explains that the principal seeks to control the

agent via performance information. The theory rests on the idea that both the principal and agent are self-interested actors and are motivated to act by this self-interest.

This dissertation supports an expanded explanation to the principal-agent theory (Moynihan et al., 2012; Gailmard & Patty, 2007; Heinrich & Marschke, 2010) by exploring other-oriented constructs and how these shape the dysfunctional use of information. Other-oriented constructs include such concepts as prosociality, public service motivation, mission-orientation, and citizenship behavior (Korsgaard et al., 1996). In my systematic review (my first essay of this dissertation), I found that individual and behavioral traits have been studied extensively in fields outside of public administration as explanations for unethicity and performance gaming. I tested these theories by examining factors embedded in other-oriented constructs, such as mission-orientation (in the observational study) and prosocial impact information (in the experimental study). I contributed to building a theory of performance gaming based on individual and behavioral traits, specifically other-oriented constructs. To the practical implications for public management, both studies ultimately address concerns of dysfunctional uses of performance information and ways public managers can mitigate the consequences of these dysfunctions. This dissertation has implications for the citizens that public managers serve because some communities (specifically historically disadvantaged groups) are affected much more by performance gaming than other groups. This chapter will provide conclusive remarks for each of my three essays. Then, I will provide a synthesized analysis and conclusion, making a summative argument that other-oriented constructs can form a basis for understanding dysfunctional uses of information.

5.2 Conclusion of Essay 1: Systematic Review

In the study of performance gaming, public administration has mainly focused on antecedents that are part of the performance system itself. These include studies that have examined the effects of system design flaws, targets, rankings, and incentives (Bevan & Hood, 2006; Courty & Marschke, 2007). In other social science fields, the focus has mainly been on individual and behavioral traits, such as moral identity and cognitive dissonance (Ayal & Gino, 2011; Welsh & Ordonez, 2014). The primary method in these other fields has been an experimental approach and the unit of analysis has been graduate students and citizens (Aquino et al., 2009). Public administration can draw lessons from these fields by looking outside of the organizational and system variables and focusing on individual and behavioral traits.

In my systematic literature review, I provided hypotheses based on my findings. The hypotheses and propositions included the need for an other-oriented theory of performance gaming, more experimental research, and the use of interactive variables. The hypotheses point to the developing a theory for explaining performance gaming that is centered on other-oriented constructs. While scholars studying private organizations may see dysfunctional use of performance information as a function of self-interest, in public organizations there is much research that points to public servants having a predisposition toward serving the public (Perry, 1996). This supports the view that other-oriented constructs may be particularly important to understanding performance gaming in the public sector. I also explained that public administration can benefit through more experimental research, which provide stronger arguments for causality. And the field can

benefit from stronger causal models that feature the use of interactive variables. In the other social science disciplines, much of the research on unethical behavior has involved experimental research and the use of interactive variables in creating more explicit models of unethicity (Bateson et al., 2006; Mazar et al., 2008; Zhong et al., 2010).

In many ways, this dissertation is a response to the findings and gaps found in the systematic review. The need for a strong theory of performance gaming was partially accomplished in this dissertation through the use of other-oriented constructs as building blocks for a theory of performance gaming. This dissertation also uses an experimental method in one of the major studies, which is a direct response to findings in my systematic review that there is a lack of experimental research in performance gaming. While laboratory tests may not always be practical for studying public managers, quasi-experimental studies and natural experiments may be suitable. Finally, this dissertation responds to the need for the use of interactive variables. It was in the interactive variable of experience and prosociality and also prosociality and attention where I made interesting findings. These interaction variables provided explanation for managerial performance gaming in my experimental study. Other-oriented constructs themselves may not hold viable explanation for behavior. This is because a single variable is seldom the lone predictor of outcomes in social science. There is a lot to be gained by the use of more sophisticated models that include interacting variables. The systematic review was a necessary guide for this dissertation as it pointed to the gaps in the performance gaming literature.

5.3 Conclusion of Essay 2: Effort Substitution

In my second essay for this dissertation, I studied effort substitution, a specific type of performance gaming where effort is substituted or shifted toward rewarded indicators at the expense of unrewarded ones. Using a five-year panel data of 64 high schools, I found that task demands can predict effort substitution. When rewarded and unrewarded indicators capture different dimensions of performance, there is strong support for the mitigating effect of mission orientation on performance gaming. The conclusion is that public managers are faced with real decisions when rewards and unrewarded indicators are not linked, so there are real decisions to make. A purpose-driven organization will have a mitigating effect on performance gaming in such situations because managers will look for purpose to guide decision-making.

This paper contributes to building a theory of performance gaming that includes other-oriented constructs as antecedents. In my paper on effort substitution, the focus is on the significance of organizations that are oriented toward their purposefully-established missions. While all organizations have missions, not all missions are purposefully-driven to serve the public and not all organizations are oriented toward their mission (Brewster & Cerdin, 2018; Knies & Leisink, 2018). In the case of public organizations, a purposefully-established mission is aimed at serving the public interest (or in some cases, specific target groups within the public sphere). This paper connects to the theme of my dissertation on the significance of other-oriented constructs but uses public organizations as the unit of analysis. It underscores the importance of organizational variables shaping the behavior of public managers. This is consistent with

the overall performance management literature that finds managerial behavior is shaped by organizational variables (Hood, 2012; Kettl, 1997; Moynihan & Pandey, 2010).

5.4 Conclusion of Essay 3: Prosocial Impact

I studied the effect of prosocial impact on cream skimming, a type of gaming where undesirable clients are excluded from service provision in order to improve performance output. Prosocial impact is an other-oriented construct that describes the experience individuals have from helping others (Bolino & Grant, 2016). I conducted two experimental studies, one using public managers and a second study using a general population sample. In this randomized experimental study, participants were placed into three experimental groups with different visualizations of prosocial impact data (one group received a placebo with no impact data). I found that prosocial impact was less influential on public managers than on the general population sample. Prosocial impact is more influential when public managers are less experienced or give more attention to performance data. The conclusion from this experimental study is that prosociality is influential in explaining performance gaming, but this effect depends on the user. For public managers who are more experienced, there is less reliance on performance data to guide decision-making and thus prosociality will have less of a mitigating effect on performance gaming. For public managers who give less attention to performance data, there will also be less reliance on this information to influence decision-making. For public managers, attention to data and experience are important intervening variables in terms of prosocial impact's effect.

My third essay also contributes to building a theory of performance gaming centered on other-oriented constructs. The study shows that other-oriented factors can explain organizational behavior rather than purely economic ones that are based on self-interest. There is also the importance of performance information itself as an antecedent of performance gaming. Different visualizations of data are processed differently and when these data are perceived to be prosocial, they can greatly influence how data are used in dysfunctional ways. I have a strong case that other-oriented constructs can explain performance gaming. In the following section, I will synthesize all three essays to make this case.

5.5 Synthesized Conclusion

The problem this dissertation addresses is the issue of why performance gaming occurs and ways to minimize it. I took a deductive approach by first theorizing on an alternative explanation for principal-agent theory, the traditional theory of public management. My theoretical explanation is based on other-oriented constructs that are added to agency theory. These other-oriented constructs are those concepts that describe how individuals help others. Other-oriented constructs are not a replacement for the principal-agent theory and do not equate to stating agency theory is a completely baseless one. Other-oriented constructs provide a different perspective for understanding how performance information is used. Though I am not the first to portend to this alternative theory (see Moynihan et al., 2012), I make the novel contribution of using this theory as a building block for understanding performance gaming. This theoretical building block is especially useful for understanding performance gaming in the context of public

organizations. This is because public administration has a long tradition of arguing that public servants have a disposition toward public service (Houston, 2000; Moynihan & Pandey, 2007; Perry & Wise, 1990). My theoretical argument and empirical findings are consistent with this tradition. Other-oriented constructs are useful for explaining organizational behavior, especially for public organizations because of the predisposition of public servants to serve the public. But my arguments are also novel because I take a behavioral perspective in studying performance information use. Rather than a systems approach to studying the dysfunctional uses of performance information, the behavioral perspective looks at the human side of motivations for organizational behavior. To make this more lucid, this dissertation's overarching finding is that the willingness to game performance systems can be mitigated through organizations being other-oriented. That is to say, organizations can focus on traits that put others first, such as aligning to their mission and being prosocial. These are traits that have been heavily studied as mitigating factors of performance gaming in other social science fields. And public administration can benefit by taking this route.

For public management practice, I highlighted the fact that managers matter in performance systems, which is consistent with the public management literature (Behn, 1995; Ingraham et al., 2003; Moynihan & Pandey, 2004). I contribute to the discussion on ways in which public managers can reduce dysfunctional uses of performance information through prosocial and mission-driven mechanisms. Most importantly, my dissertation contributes to a discussion on how dysfunctional uses of performance information affect public service. The fact that my research had mixed findings on the role historically disadvantaged groups play in dysfunctional uses of information warrants

a discussion here and further research. In this section, I synthesize my research to make the case for other-oriented constructs as antecedents of performance gaming. I conclude this section with my overall theoretical and practical contributions and why my findings matter for public service.

5.5.1 Other-Oriented Constructs: Toward a Theory of Performance Gaming

This dissertation studies the effects of prosociality and mission-orientation as other oriented constructs and builds toward a theory of performance gaming based on these other-oriented constructs. Such constructs describe individuals' propensity to help others. And they enhance the principal-agent theory that has traditionally explained how agents (public managers) behave toward principals (elected officials) and the role of performance information in this relationship. The traditional agency theory states that performance information is the mechanism by which principals control agents (Jensen & Meckling, 1976). In performance systems, agents report information to principals who determine whether the agents are meeting the goals and executing policies set forth by principals. In this relationship, agents (as well as principals) are motivated by self-interest, which has been used to explain why agents use performance information. In public administration, some scholars have examined the role of altruistic motivations to explain why performance information is used in purposeful ways (Moynihan, et al., 2012). Previous research has implied there is an alternative explanation of agency theory (Gailmard & Patty, 2007; Heinrich & Marschke, 2010). My dissertation adds to this alternative view of agency theory, but tests the theory on performance gaming, a dysfunctional use of performance information. Also, my dissertation tests this theory in the context of public organizations. I find support for other-oriented constructs (mission

orientation and prosociality) as explanations for dysfunctional behavior in public organization. For mission-orientation, the argument is based on an organizational perspective and for prosociality, the argument is supported in an individual context.

Mission orientation is the other-oriented construct that is associated with reducing performance gaming for organizations. While all organizations have a mission, I make a distinction to the type of mission orientation I study in this dissertation. In private organizations, an organization's mission is usually tied to profits or expanding the organization's influence. This is a mission based on organizational self-interest. For public organizations, the mission is usually tied to serving the public interest or a social cause. This latter concept of mission is what comprises the mission orientation that I study in this dissertation. This is what makes mission orientation an other-oriented construct. It is difficult to measure the degree to which organizations rely on their mission, but most studies that have examined the effects of missions on organizational behavior have used mission statements. In this dissertation, I use mission statements of public organizations to measure mission orientation, as well as specific types of mission-driven organizations (charter schools). Organizations with missions that emphasize specific groups or special purposes are differentiated from organizations with general missions. So, my dissertation establishes the importance of mission orientation as an other-oriented construct for organizations. Through my five-year longitudinal study, my dissertation also provides empirical support for mission orientation as an other-oriented predictor of performance gaming. Although, my findings were mixed for mission orientation as a direct effect on performance gaming, mission orientation is more profound as a predictor under specific conditions in the performance system. For

example, mission orientation predicts performance gaming when unrewarded and rewarded indicators are not conceptually linked. So, organizations with a strong mission orientation can expect to reduce gaming behavior. But more importantly, there is evidence toward a strong theory of performance gaming that is based on organizations that are other-oriented.

From the individual perspective, I study prosocial impact as an other-oriented construct that reduces performance gaming. Prosocial impact occurs when individuals experience others benefiting from their work. This dissertation examines prosocial impact framed through different visualizations of performance information. The argument that prosocial impact can have benefits for performance systems is supported by the literature in prosociality (Bolino & Grant, 2016). Individuals can become more motivated for work and they can improve their performance because of prosocial factors. Like mission orientation, prosociality fits within the tradition of public administration scholarship that finds public servants are predisposed toward service to the public. But prosociality is a broader term because it encompasses not just serving the public but anyone other than self (Esteve, Urbig, van Witteloostuijn, & Boyne, 2016; Perry, Hondeghem, & Wise, 2010). Prosociality can include helping fellow co-workers, supervisors, or even competing agencies. By framing prosocial impact within the performance information itself and also measuring prosocial impact as individuals' perceptions, I was able to highlight nuances in prosociality. Prosocial impact was a significant predictor when it was an intervening variable, such as with the public manager sample where experience and time intervened with prosocial impact. From my experimental study, public managers differ from the general population in how they process prosocial impact information.

Prosocial impact information that is qualitative and highlights historically disadvantaged groups (such as African Americans) was more impactful for the general population than for public managers. From this dissertation, prosociality, as an other-oriented construct, may be different for different groups. Public managers may be less influenced by prosocial impact information when it comes to performance gaming, but this can be due to the visualization of the information itself. Public managers may not be affected by visualization cues in the same way as the general population.

To conclude this section, I have evidence toward a theory of performance gaming that is based on other-oriented constructs. These constructs describe how organizations and individuals are disposed toward helping others. From the organizational perspective, there is mission orientation, which describes organizations that are aligned to a specific purpose in service to others. From the individual perspective, there is prosocial impact, where individuals experience how their work benefits others. While my empirical studies provide mixed support for direct effects of other-oriented constructs in the public management sector, there are nuances that my research has uncovered. For example, when other-oriented constructs work through intervening variables, their effect is significant. In the case of mission orientation, when unrewarded and rewarded indicators are conceptually linked, mission orientation has a significant effect on performance gaming. This dissertation establishes a concrete building block for a theory of performance gaming where other-oriented constructs mitigate performance gaming. In the next section, I describe this theoretical contribution among others as well as the significance for public management practice.

5.5.2 Contributions to theory and practice

This dissertation makes a contribution to management theory in the area of performance management by researching performance gaming. I provide a building block for a theory of performance gaming that highlights other-oriented constructs as factors that reduce gaming. This dissertation also makes a contribution to management practice with implications for managerial decision-making and improving organizational effectiveness. Although I studied public organizations in this dissertation, I made efforts for a cross-disciplinary connection through my systematic review. Using lessons from this review, I expand on some of the research taking place in these other disciplines, with emphasis on human motivational constructs that can explain dysfunctional behavior. Finally, for the public service, this dissertation has practical significance because it leads to a discussion of improving service delivery, especially among historically disadvantaged groups, which are severely affected by performance gaming. In this section, I discuss these points in more detail.

First, performance management does not have a specific theory that explains why performance gaming occurs. To understand the behavior of public managers in a performance system, our field has relied on the concept of self-interest as explained in principal-agent theory. To explain organizational behavior in performance systems, we have often studied system norms and cultures that explain why firms behave the way they do (Bevan & Hood, 2006; Courty & Marschke, 2007; de Bruijn, 2002). In my dissertation, using an alternative explanation of principal-agent theory, in line with other studies in performance management (Gailmard & Patty, 2007; Moynihan et al., 2012), I have turned to other-oriented constructs. These constructs rely on altruistic motivations to

explain behavior. Also, the role of behavioral and human motivational variables is encouraged by the other social science disciplines (Ayal et al., 2015; Barkan, et al., 2012). In these disciplines, self-image and moral justification are studied as antecedents of dysfunctional behavior. Therefore, my dissertation contributes to the literature by explaining performance gaming through the prism of these other-oriented constructs. Mission orientation and prosocial impact do affect managerial attitudes toward performance gaming. For example, when faced with conceptually different dimensions of performance, public managers will be more susceptible to mission orientation and depend more on their purpose and mission in making decisions. Managers can use mission statements and moral reminders as cues for guiding their organizations toward the mission.

Second, for management practice, my dissertation contributes to approaches for management that can improve decision-making and organizational effectiveness. Public managers are faced with tough choices and extreme pressures in performance systems. Under these conditions, managers might be tempted to cut corners or fudge performance numbers and engage in other dysfunctional behavior. However, managers that turn to their organization's mission or are reminded of that purpose, may depend more on altruistic motivations that mitigate a desire to undermine their organization's goal. So, one recommendation for practice is that organizations remind individuals of their purpose and mission and ensure that functions are aligned to these purposes. This can include managers using moral cues (goals and purpose charts displayed in the work environment), mission statement emphasis, and meetings with employees that emphasize the organization's purpose. Also, for decision-making and improved effectiveness,

managers can spend more time on performance information and ensure that more information is prosocial. This can lead to better decision-making and thus a more effective organization. In sum, public organizations can become more effective when they can connect managers to these altruistic factors – the organization’s purpose and to spend more time on the prosocial impact of information.

Third, my dissertation encourages a discussion on performance gaming and its effect on the general public. This connection to public service is a primary reason why this dissertation is unique to public organizations. In the public sector, performance gaming not only makes organizations ineffective because it undermines organizational goals, but it also harms the particular group that the public organization serves (not to mention it undermines public trust and democratic values). Although the public as a whole is harmed, certain groups become doubly disadvantaged as a result of dysfunctional behaviors. In my dissertation, I made the connection between task demands and effort substitution. From a policy perspective, task demands in the public sector are associated with low-income individuals, high-crime areas, people with severe disabilities, and racial and ethnic groups, such as Blacks or Hispanics. This translates to these groups being the most common recipients of performance gaming’s negative effects. In my experimental study, prosocial impact information is more influential in reducing performance gaming when beneficiaries are from disadvantaged groups (Black students). However, this is only true for the general population sample. There is no effect for public managers. The discussion here is that public managers and the general public may perceive groups differently. While this dissertation does not uncover the reasons for the difference, it opens up a conversation on how the public may perceive these groups

differently. The role that particular groups play in performance gaming is still left unanswered. But this is an important avenue for further research.

In conclusion, my dissertation establishes the importance of other-oriented constructs in understanding performance gaming. Prosociality and mission orientation are just two of these constructs studied here. But there are- moderating variables, such as attention to information and managerial experience. Individual traits play an important role in dysfunctional behavior as studied in other disciplines. My dissertation brings these specific traits to the forefront of public administration research in performance gaming. Public managers can benefit from this research by creating organizations that are more purpose-driven or aligned with existing missions. They can give more attention to information that is prosocial. For public service, performance gaming is a serious consequence of performance systems that can harm the general public in how they receive government services. We still do not know to what extent particular groups are harmed or helped by other-oriented constructs as part of the performance gaming model presented in this dissertation. However, we do know that historically disadvantaged groups are often doubly disadvantaged by performance gaming and so future research should examine this dynamic. As research on performance gaming grows, it is my hope that we can study more individual traits and build toward a specific theory of performance gaming.

REFERENCES

- Abeler, J., Becker, A., & Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economics*, 113, 96–104.
- Abelman, R., Atkin, D., Dalessandro, A., Snyder-Suhy, S., & Janstova, P. (2007). The trickle down effect of institutional vision: Vision statements and academic advising. *NACADA Journal*, 27(1), 4-21.
- Alge, B. J., Greenberg, J., & Brinsfield, C. T. (2006). An identity-based model of organizational monitoring: Integrating information privacy and organizational justice. *Research in Personnel and Human Resources Management*, 25, 71-135.
- Alon-Barkat, S., & Gilad, S. (2017). Compensating for poor performance with promotional symbols: Evidence from a survey experiment. *Journal of Public Administration Research and Theory*, 27(4), 661-675.
- Altonji, J. G., Huang, C. I., & Taber, C. R. (2015). Estimating the cream skimming effect of school choice. *Journal of Political Economy*, 123(2), 266-324.
- Andersen, S. C., & Moynihan, D. P. (2016). How leaders respond to diversity: The moderating role of organizational culture on performance information use. *Journal of Public Administration Research and Theory*, 26(3), 448-460.
- Andrade, E.B., & Ho, T.H. (2009). Gaming Emotions in Social Interactions. *Journal of Consumer Research*, 36(4), 539-552.
- Andrews, R. (2010). Organizational social capital, structure and performance. *Human Relations*, 63(5), 583-608.
- Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83(6), 1423.
- Aquino, K., Freeman, D., Reed, A., Felps, W. and Lim, V.K. (2009). Testing a social cognitive model of moral behavior: The interactive influence of situations and moral identity centrality. *Journal of Personality and Social Psychology*, 97(1), 123-141.
- Ashforth, B.E., & Anand, V. (2003). The normalization of corruption in organizations. *Research in Organizational Behavior*, 25, 1–52.
- Ayal, S., Gino, F., Barkan, R., & Ariely, D. (2015). Three principles to REVISE people’s unethical behavior. *Perspectives on Psychological Science*, 10(6), 738-741.
- Augoustinos, M., & Walker, I. (1996). *Social cognition: An integrated introduction*. Sage.

- Baird, K.M., & Harrison, G.L. (2017). The association between organizational culture and the use of management initiatives in the public sector. *Financial Accountability & Management*, 33(3), 311-329.
- Banerjee, R., Gupta, N. D., & Villeval, M. C. (2018). The spillover effects of affirmative action on competitiveness and unethical behavior. *European Economic Review*, 101, 567-604.
- Barkan, R., Ayal, S., Gino, F. and Ariely, D. (2012), “The pot calling the kettle black: distancing response to ethical dissonance”, *Journal of Experimental Psychology*, Vol. 141, pp. 757–773.
- Barnett, M. (2008). An attention-based view of real options reasoning. *Academy of Management Review* 33(3), 606–628.
- Baumeister, R. F. (1998). The self. In D.T. Gilbert, S.T. Fiske, & G. Lindzey (Eds.), *Handbook of Social Psychology*, (Vol. 1, pp. 680-740). McGraw-Hill.
- Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2), 169.
- Belle, N. (2013a). Experimental evidence on the relationship between public service motivation and job performance. *Public Administration Review*, 73, 143–153.
- _____. (2013b). Leading to make a difference: A field experiment on the performance effects of transformational leadership, perceived social impact, and public service motivation. *Journal of Public Administration Research and Theory*, 24(1), 109-136.
- Belle, N., & Cantarelli, P. (2017a). What causes unethical behavior? A meta-analysis to set an agenda for public administration research. *Public Administration Review*, 77(3), 327–393.
- _____. (2019b). Do ethical leadership, visibility, external regulation, and prosocial impact affect unethical behavior? Evidence from a laboratory and a field experiment. *Review of Public Personnel Administration*, 39(3), 349-371.
- Belot, M., & Schroder, M. (2013). Sloppy work, lies and theft: A novel experimental design to study counterproductive behavior. *Journal of Economic Behavior and Organization*, 93, 233–238.
- Benabou, R., & Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. *Quarterly Journal of Economics*, 126(2), 805–855.

- Benaine, S. L., & Kroll, A. (2019). Explaining effort substitution in performance systems: The role of task demands and mission orientation. *Public Management Review*, 1-23.
- Bevan, G., and Hamblin, R. (2009). Hitting and missing targets by ambulance services for emergency calls: Effects of different systems of performance measurement within the UK. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(1), 161-190.
- Bevan, G. and Hood, C. (2006). What's measured is what matters: targets and gaming in the English Public Health Care System", *Public Administration*, Vol. 84 No. 3, pp. 517-538.
- Binning, J. J. L. (2009). Coherent conceptualization is useful for many things, and understanding validity is one of them. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2(4), 486-492.
- Blinder, A. (2015, April 1). Atlanta educators convicted in school cheating scandal. *The New York Times*. <https://www.nytimes.com/2015/04/02/us/verdict-reached-in-atlanta-school-testing-trial.html>
- Bloodgood, J., Turnley, W.H., & Mudrack, P.E. (2010). Ethics instruction and the perceived acceptability of cheating. *Journal of Business Ethics*, 95(1), 23-37.
- Bohte, J. and Meier, K. (2000), "Goal displacement: assessing the motivation for organizational cheating", *Public Administration Review*, Vol. 60 No. 2, pp. 173-82.
- Bolino, M. and Grant, A.M. (2016), "The bright side of being prosocial at work, and the dark side, too: a review and agenda for research on other-oriented motives, behavior, and impact in organizations", *The Academy of Management Annals*, Vol. 10 No. 1, pp. 599-670.
- Boerema, A. (2006). An analysis of private school mission statements. *Peabody Journal of Education*, 81(1), 180-202.
- Bozeman, B., & Johnson, J. (2015). The political economy of public values: A case for the public sphere and progressive opportunity. *The American Review of Public Administration*, 45(1), 61-85.
- Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). Explaining the short careers of high-achievement teachers in schools with low-performing students. *American Economic Review*, 95(2), 166-171.

- Boyne, G. A., & Chen, A. A. (2007). Performance targets and public service improvement. *Journal of Public Administration Research and Theory*, 17(3), 455-477.
- Brewster, C., & Cerdin, J. L. (2018). The management of people in mission-driven organizations. In C. Brewster, & J. L. Cerdin (Eds.), *HRM in mission driven organizations: Managing people in the not for profit sector* (pp.1-14). Palgrave Macmillan.
- Brown, H., Henig, J., Lacireno-Paquet, N., & Holyoke, T. T. (2004). Scale of operations and locus of control in market-versus mission-oriented charter schools. *Social Science Quarterly*, 85(5), 1035-1051.
- Burgess, S., & Ratto, M. (2003). The role of incentives in the public sector: Issues and Evidence. *Oxford Review of Economic Policy*, 19(2), 285-300.
- Burton, B. & Near, J. (1995). Estimating the incidence of wrongdoing and whistle blowing: Results of a study using randomized response technique. *Journal of Business Ethics*, 14(1), 17-30.
- Cadsby, C.B., Du, N., & Song, F. (2016). In-group favoritism and moral decision making. *Journal of Economic Behavior & Organization*, 128, 59–71.
- Cannata, M. (2007). Teacher community in elementary charter schools. *Education Policy Analysis Archives*, 15(11), 1-31.
- Carhart, M., Kaniel, R., Musto, D.K., & Reed, A.V. (2002). Leaning for the tape: Evidence of gaming behavior in equity mutual funds. *The Journal of Finance*, 57(2), 661-693.
- Cepiku, D., Hinna, A., Scarozza, D., & Savignon, A.B. (2017). Performance information use in public administration: An exploratory study of determinants and effects. *Journal of Management & Governance*, 21(4), 963-991.
- Chun, Y. H., & Rainey, H. G. (2005). Goal ambiguity and organizational performance in U.S. federal agencies. *Journal of Public Administration Research and Theory*, 15(4), 529-557.
- Cohen, E., Franck, V., Kelliher, K., Varghese, P., & Walsh, K. (2012). *Teacher quality roadmap: Improving policies and practices in the Miami Dade County Public Schools*. National Council on Teacher Quality, Washington, D.C.
https://www.nctq.org/dmsView/Teacher_Quality_Roadmap_Improving_Policies_and_Practices_in_Miami_NCTQ_Report

- Cohn, A., Maréchal, M., & Noll, T. (2013). Saliency of criminal identity causes dishonest behavior: An experiment behind bars. *University of Zurich Manuscript*, 1-29
- Cojoc, D.C., & Stoian, A. (2014). Dishonesty and charitable behavior. *Experimental Economics*, 17, 717–732.
- Coppock, A. (2019). Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. *Political Science Research and Methods*, 7(3), 613-628.
- Courty, P., Heinrich, C., & Marschke, G. (2005). Setting the standard in performance measurement systems. *International Public Management Journal*, 8(3), 321-347.
- Courty, P., & Marschke, G. (2003). Dynamics of performance-measurement systems. *Oxford Review of Economic Policy*, 19(2), 268–284.
- _____. (2007), “Making government accountable: lessons from a federal job training program”, *Public Administration Review*, Vol. 67 No. 5, pp. 904-916.
- Cropanzano, R., & Mitchell, M. S. (2005). Social exchange theory: An interdisciplinary review. *Journal of Management*, 31(6), 874-900.
- Crossan, M., Mazutis, D., & Seijts, G. (2013). In search of virtue: The role of virtues, values and character strengths in ethical decision making. *Journal of Business Ethics*, 113(4), 567-581.
- Cuganesan, S., Guthrie, J., & Vranic, V. (2014). The riskiness of public sector performance measurement: A review and research agenda. *Financial Accountability & Management*, 30(3), 279-302.
- Desmidt, S. (2016). The relevance of mission statements: Analysing the antecedents of perceived message quality and its relationship to employee mission engagement. *Public Management Review*, 18(6), 894-917.
- Destler, K.N. (2016). Creating a performance culture: incentives, climate, and organizational change. *The American Review of Public Administration*, 46(2), 201-225.
- Doolittle, A., & Faul, A. C. (2013). Civic engagement scale: A validation study. *Sage Open*, 3(3), 1-7.

- Douglas, S., Hood, C., Overmans, T., & Scheepers, F. (2019). Gaming the system: Building an online management game to spread and gather insights into the dynamics of performance management systems”, *Public Management Review*, 21(10), 1560-1576.
- Dynarski, S. (2018, September 28). At elite colleges, racial diversity requires affirmative action. *The New York Times*. <https://www.nytimes.com/2018/09/28/business/at-elite-colleges-racial-diversity-requires-affirmative-action.html>.
- Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, 14(1), 57-74
- Emma, C. (2015, November 3). Here’s why \$7 billion didn’t help America’s worst schools. *Politico*. <https://www.politico.com/story/2015/11/failing-schools-education-white-house-214332>.
- Erat, S., & Gneezy, U. (2012). White lies. *Management Science*, 58(4), 723-733.
- Esteve, M., Urbig, D., Van Witteloostuijn, A., & Boyne, G. (2016). Prosocial behavior and public service motivation. *Public Administration Review*, 76(1), 177-187.
- Eterno, J., & Silverman, E. (2012). *The crime numbers game. Management by manipulation*. CRC Press.
- Faravelli, M., Friesena, L., and Gangadharan, L. (2015). Selection, tournaments, and dishonesty. *Journal of Economic Behavior & Organization*, 110, 160-175.
- Filiz, E., & Battaglio Jr, R. P. (2017). Personality and decision-making in public administration: the five-factor model in cultural perspective. *International Review of Administrative Sciences*, 83(1_suppl), 3-22.
- Fischbacher, U. and Föllmi-Heusi, F. (2013), “Lies in disguise-an experimental study on cheating”, *Journal of the European Economic Association*, Vol. 11 No. 3, pp. 525-547.
- Florida Department of Education. (2016). *Florida’s high school cohort 2015-16 graduation rate*. <http://www.fldoe.org/core/fileparse.php/7584/urlt/GradRates1516.pdf>.
- Florida Department of Education. 2017. *2016-17 Guide to calculating school and district grades*. <http://www.fldoe.org/core/fileparse.php/18534/urlt/SchoolGradesCalcGuide17.pdf>.

- Florida Department of Education. (2018). *Florida School Recognition Program*.
<http://www.fldoe.org/accountability/accountability-reporting/fl-school-recognition-program/>
- Ford, M. R., & Ihrke, D. M. (2016). Comparing nonprofit charter and traditional public school board member perceptions of the public, conflict, and financial responsibility: Is there a difference and does it matter? *Public Management Review*, 18(7), 972-992.
- Frey, B.S., Homberg, F., & Osterloh, M. (2013). Organizational control systems and pay for performance in the public service. *Organization Studies*, 34(7), 949-972.
- Gailmard, S., & Patty, J. W. (2007). Slackers and zealots: Civil service, policy discretion, and bureaucratic expertise. *American Journal of Political Science*, 51(4), 873–889.
- Garrow, E. E., & Grusky, O. (2013). Institutional logic and street-level discretion: The case of HIV test counseling. *Journal of Public Administration Research and Theory*, 23(1), 103–131.
- Gay, C. (2006). Seeing Difference: The effect of economic disparity on Black attitudes toward Latinos. *American Journal of Political Science*, 50(4), 982-997.
- Gerrish, Ed. (2016). The impact of performance management on performance in public organizations: A meta-analysis. *Public Administration Review*, 76(1), 48-66.
- Gibbs, M.J., Merchant, K.A., Van der Stede, W.A., & Vargus, M.E. (2009). Performance measure properties and incentive system design. *Industrial Relations: A Journal of Economy and Society*, 48(2), 237-264.
- Gino, F. (2015). Understanding ordinary unethical behavior: Why people who value morality act immorally. *Current Opinion in Behavioral Sciences* 3, 107–111.
- Gino, F., Ayal, S. and Ariely, D. (2009), “Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel”, *Psychological Science*, Vol. 20 No. 3, pp. 393-398.
- Gino, F., Ayal, S., & Ariely, D. (2013). Self-serving altruism? The lure of unethical Actions that benefit others. *Journal of Economic Behavior & Organization*, 93, 285-292.
- Glac, K., Warren, D.E., & Chen, C.C. (2012). Conflict in roles: Lying to the in-group versus the out-group in negotiations. *Business & Society*, 53(3), 440–460.

- Goodsell, C. (1994). *The case for bureaucracy*. Chatham House.
- Grant, A. M. (2008a). Employees without a cause: the motivational effects of prosocial impact in public service. *International Public Management Journal*, 11, 48–66.
- . (2008b). The significance of task significance: Job performance effects, relational mechanisms, and boundary conditions. *Journal of Applied Psychology*, 93(1), 108-124.
- . (2008c). Does intrinsic motivation fuel the prosocial fire? Motivational synergy in predicting persistence, performance, and productivity. *Journal of Applied Psychology*, 93, 48–58.
- . (2012). Leading with meaning: beneficiary contact, prosocial impact, and the performance effects of transformational leadership. *Academy of Management Journal*, 55(2), 458–476.
- Grant, A.M., & Campbell, E.M. (2007). Doing good, doing harm, being well and burning out: the interactions of perceived prosocial and antisocial impact in service work. *Journal of Occupational and Organizational Psychology*, 82(4), 665-691.
- Grimes, P.W., & Rezek, J.P. (2005). The determinants of cheating by high school economics students: A comparative study of academic dishonesty in the transitional economies. *International Review of Economics Education*, 4(2), 23-45.
- Grolleau, G., Kocher, M.G., & Sutan, A. (2016). Cheating and loss aversion: Do people cheat more to avoid a loss? *Management Science*, 62(12), 3428-3438.
- Grøn, C. H., Bro, L. L., & Andersen, L. B. (2019). Public managers' leadership identity: Concept, causes, and consequences. *Public Management Review*, 1-21.
- Hannah, S.T., Avolio, B.J., & May, D.R. (2011). Moral maturation and moral conation: a capacity approach to explaining moral thought and action. *Academy of Management Review*, 36(4), 663-685.
- Hassan, S., Wright, B. E., & Yukl, G. (2014). Does ethical leadership matter in government? Effects on organizational commitment, absenteeism, and willingness to report ethical problems. *Public Administration Review*, 74(3), 333-343.
- Hasson, Y., Tamir, M., Brahm, K. S., Cohrs, J. C., & Halperin, E. (2018). Are liberals and conservatives equally motivated to feel empathy toward others? *Personality and Social Psychology Bulletin*, 44(10), 1449-1459.

- Hatry, H. (2007). *Performance measurement: Getting results*. Urban Institute Press.
- Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.
- Heath, G., & Radcliffe, J. (2007). Performance measurement and the English Ambulance Service. *Public Money and Management*, 27(3), 223-228.
- Heckman, J., Heinrich, C., & Smith, J. (2002). The performance of performance standards. *The Journal of Human Resources*, 37(4), 778-811.
- Heckman, J. J., Heinrich, C. J., Courty, P., Marschke, G., & Smith, J. (2011). *The performance of performance standards*. W. E. Upjohn Institute for Employment Research. https://research.upjohn.org/up_press/209/
- Heinrich, C. J. (2007). False or fitting recognition? The use of high performance bonuses in motivating organizational achievements. *Journal of Policy Analysis and Management*. 26(2) 281–304.
- Heinrich, C., & Marschke, G. (2010). Incentives and their dynamics in public sector performance management systems. *Journal of Policy Analysis and Management*, 29(1), 183-208.
- Henig, J. R., Holyoke, T. T., Brown, H., & Lacireno-Paquet, N. (2005). The influence of founder type on charter school structures and operations. *American Journal of Education*, 111(4), 487-588.
- Hershey, D. A., & Walsh, D. A. (2000). Knowledge versus experience in financial problem solving performance. *Current Psychology*, 19(4), 261-291.
- Hertz, S.G., & Krettenauer, T. (2016). Does moral identity effectively predict moral behavior?: A meta-analysis. *Review of General Psychology*, 20(2), 129-140.
- Hilbig, B.E., & Hessler, C.M. (2013). What lies beneath: How the distance between truth and lie drives dishonesty. *Journal of Experimental Social Psychology*, 49(2), 263–266.
- Hill, C. J., & Lynn Jr, L. E. (2015). *Public management: Thinking and acting in three dimensions*. CQ Press.
- Hill, J., & Kochendorfer, R.A. (1969). Knowledge of peer success and risk of detection as determinants of cheating. *Developmental Psychology*, 1(3), 231–238.

- Hine, K. A., Porter, L. E., Westera, N. J., Alpert, G. P., & Allen, A. (2018). Exploring police use of force decision-making processes and impairments using a naturalistic decision-making approach. *Criminal Justice and Behavior*, 45(11), 1782-1801.
- Hood, C. (2006), "Gaming in targetworld: the targets approach to managing British public services", *Public Administration Review*, Vol. 66 No. 4, pp. 515-521.
- . (2007). Public service management by numbers: Why does it vary? Where has it come from? What are the gaps and the puzzles? *Public Money and Management*, 27(2), 95-102.
- . (2012). Public management by numbers as a performance-enhancing drug: two hypotheses. *Public Administration Review*, 72(S1), S85-S92.
- Hood, C., & Dixon, R. (2010). The political payoff from performance target systems: No brainer or no-gainer? *Journal of Public Administration Research and Theory*, 20(2), i281-i298.
- Howard, C., Freeman, A., Wilson, A., & Brown, E. (2017). Poverty. *Public Opinion Quarterly*, 81(3), 769-789.
- Hruschka, D., Efferson, C., Jiang, T., Falletta-Cowden, A., Sigurdsson, S., McNamara, R., Sands, M., Munira, S., Slingerland, E., & Henrich, J. (2014). Impartial institutions, pathogen stress and the expanding social network. *Human Nature*, 25(4), 567-579.
- Hu, J., & Liden, R. C. (2015). Making a difference in the teamwork: Linking team prosocial motivation to team processes and effectiveness. *Academy of Management Journal*, 58(4), 1102–1127.
- Huff, C., & Tingley, D. (2015). Who are these people? Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*, 2(3).
- Hurkens, S., & Kartik, N. (2009). Would I lie to you? On social preferences and lying Aversion. *Experimental Economics*, 12(2), 180–192.
- Hyman, D.A. (2001). Health care fraud and abuse: Market change, social norms, and the trust 'reposed in the workmen'. *The Journal of Legal Studies*, 30(S2), 531-567.

- Iasevoli, B. (2016, November 23). Teacher bonuses are paying off in Broward County. *Education Week*.
http://blogs.edweek.org/edweek/teacherbeat/2016/11/teacher_bonuses_are_paying_off.html.
- Ibañez, M., & Riener, G. (2018). Sorting through affirmative action: Three field experiments in Colombia. *Journal of Labor Economics*, 36(2), 437-478.
- Iyer, R., Koleva, S., Graham, J., Ditto, P., & Haidt, J. (2012). Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PLoS ONE*, 7(8), 1-23.
- Jacob, Brian, and Steven Levitt. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics*, 118(3), 843-877.
- Jacobsen, C., Fosgaard, T.R., & Pascual-Ezama, D. (2018). Why do we lie? A practical guide to the dishonesty literature. *Journal of Economic Surveys*, 32(2), 357–387.
- Jakobsen, M.L., Baekgaard, M., Moynihan, D.P., & van Loon, N. (2017). Making sense of Performance Regimes: Rebalancing external accountability and internal learning. *Perspectives on Public Management and Governance*, 2(1), 127-141.
- Jamison, P. (2018, March 8). D.C. Public Schools scandals: A short guide. *The Washington Post*. https://www.washingtonpost.com/local/dc-politics/a-short-guide-to-dc-public-schools-scandals/2018/03/08/c40e2c4e-2170-11e8-94da-ebf9d112159c_story.html
- Janis, I. L. (1971). Groupthink. *Psychology Today*, 5(6), 43-46.
- Jennings, E.T., & Haist, M.P. (2004). Putting performance measurement in context. In P.W. Ingraham, & L.E. Lynn Jr., (Eds.), *The art of governance: Analyzing management and administration* (pp.173-194). Georgetown University Press.
- Jensen, M., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency cost, and ownership structure. *Journal of Financial Economics*, 3(4), 305-360.
- Jilke, S., & Tummers, L. (2018). Which clients are deserving of help? A theoretical model and experimental test. *Journal of Public Administration Research and Theory*, 28(2), 226-238.
- Jones, T. (1991). Ethical decision making by individuals in organizations: An issue contingent model. *The Academy of Management Review*, 16(2), 366-395.

- Jones, L.R., & Euske, K.J. (1991). Strategic misrepresentation in budgeting. *Journal of Public Administration Research and Theory*, 1(4), 437-460.
- Jorgensen, T. B., & Bozeman, B. (2007). Public values: An inventory. *Administration & Society*, 39(3): 360-361.
- Juenke, E.G. (2005). Management tenure and network time: How experience affects bureaucratic dynamics. *Journal of Public Administration Research and Theory*, 15(1), 113-131.
- Kahan, D. M. (2015). The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it. *Emerging Trends in the Social and Behavioral Sciences*, 1-16.
- Kasperkevic, J. (2015, April 1). Georgia cheating scandal: 11 teachers found guilty of racketeering. *The Guardian*.
<https://www.theguardian.com/us-news/2015/apr/01/atlanta-teachers-found-guilty-cheating>
- Kaufmann, J.B., West, T., Ravenscroft, S.P., & Shrader, C.B. (2005). Ethical distancing: Rationalizing violations of organizational norms. *Business & Professional Ethics Journal*, 24(3), 101-134.
- Kelman, S., and Friedman, J. (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*, Vol. 19 No. 4, pp. 917-946.
- Kerpershoek, E., Groenleer, M., & de Bruijn, H. (2016). Unintended responses to performance management in Dutch hospital care: Bringing together the managerial and professional perspectives. *Public Management Review*, 18(3), 417-436.
- Kettl, D. F. (1997). The global revolution in public management: Driving themes, missing links. *Journal of Policy Analysis and Management*, 16(3), 446–462.
- Knol, M. J., Groenwold, R. H. H., & Grobbee, D. E. (2012). P-Values in baseline tables of randomised controlled trials are inappropriate but still common in high impact journals. *European Journal of Preventive Cardiology*, 19(2), 231–232.
- Koning, P., & Heinrich, C. J. (2013). Cream-skimming, parking and other intended and unintended effects of high-powered, performance-based contracts. *Journal of Policy Analysis & Management*, 32(3), 461-483.

- Korsgaard, M. A., Meglino, B. M., Lester, S. W., & Jeong, S. S. (2010). Paying you back or paying me forward: Understanding rewarded and unrewarded organizational citizenship behavior. *Journal of Applied Psychology, 95*(2), 277-290.
- Kroll, A. (2013). The other type of performance information: Nonroutine feedback, its relevance and use. *Public Administration Review, 73*(2), 265–276.
- _____. (2015). Drivers of performance information use: Systematic literature review and directions for future research. *Public Performance and Management Review, 38*(3), 459-486.
- _____. (2017). Can performance management foster social equity? Stakeholder power, protective institutions, and minority representation. *Public Administration, 95*(1), 22-38.
- Larkin, I. (2014). The cost of high-powered incentives: Employee gaming in enterprise software sales. *Journal of Labor Economics, 32*(2), 199-227.
- Le Grand, J. (2006). *Motivation, agency, and public policy: Of knights and knaves*. Oxford University Press
- Lerner, M. (2015). School quality has a mighty influence on neighborhood choice, home values. *The Washington Post*. https://www.washingtonpost.com/realestate/school-quality-has-a-mighty-influence-on-neighborhood-choice-home-values/2015/09/03/826c289a-46ad-11e5-8ab4-c73967a143d3_story.html?utm_term=.a907ada750c2.
- Lewis, J. M., & Triantafillou, P. (2012). From performance measurement to learning: A new source of government overload? *International Review of Administrative Sciences, 78*(4), 597-614.
- Li, J. (2015). The paradox of performance regimes: strategic responses to target regimes in Chinese local government. *Public Administration, 93*(4), 1152–1167.
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of Internal Medicine, 151*(4), W-65.

- Lucas, G. J., Marius, H. M., Zijlmans, M. T., & Blettner, D. P. (2017). The effect of organizational performance feedback on team attention focus. In K. J. Sund, R. J. Galavan, & A. S. Huff (Eds.), *Uncertainty and strategic decision making* (pp. 171-190). Emerald Group Publishing Limited.
- Manaf, H. A., Armstrong, S. J., Lawton, A., & Harvey, W. S. (2018). Managerial tacit knowledge, individual performance, and the moderating role of employee personality. *International Journal of Public Administration*, 41(15), 1258-1270.
- Marshall, C., & Rossman, G. B. (2016). *Designing qualitative research* (5th edition). Sage Publications.
- Marvel, J. D. (2015). Unconscious bias in citizens' evaluations of public sector performance. *Journal of Public Administration Research and Theory*, 26(1), 143-158.
- Maynard-Moody, S., & Musheno, M. (2012). Social equities and inequities in practice: street-level workers as agents and pragmatists. *Public Administration Review*, 72(s1), S16-S23.
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self concept maintenance. *Journal of Marketing Research*, 45(6), 633-644.
- McCabe, D., & Trevino, L.K. (1993). Academic dishonesty: Honor codes and other contextual influences. *Journal of Higher Education*, 64(5), 522-538.
- McGowan, R. P., & Loveless, S. (1981). Strategies for information management: The administrator's perspective. *Public Administration Review*, 41(3), 331-339.
- Mead, N.L., Baumeister, R.F., Gino, F., Schweitzer, M.E., & Ariely, D. (2009). Too tired to tell the truth: self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology*, 45(3), 594-597.
- Meanwell, E., & Swando, J. (2013). Who deserves good schools? Cultural categories of worth and school finance reform. *Sociological Perspectives*, 56(4), 495-522.
- Meier, K. J. (2016). *The politics of sin: Drugs, alcohol and public policy*. Routledge.
- _____. (2019). Theoretical frontiers in representative bureaucracy: New directions for research. *Perspectives on Public Management and Governance*, 2(1), 39-56.

- Meier, K., & Krause, G. (2003). Conclusion: An agenda for the scientific study of Bureaucracy. In K. Meier, & G. Krause (Eds.), *Politics, policy, and organizations: frontiers in the scientific study of bureaucracy* (pp. 292-308). University of Michigan Press.
- Meier, K. J., & O'Toole Jr, L. J. (2009). The proverbs of new public management: Lessons from an evidence-based research agenda. *The American Review of Public Administration*, 39(1), 4-22.
- Meier, K.J., O'Toole Jr, L.J., Boyne, G.A., & Walker, R.M. (2007). Strategic management and the performance of public organizations: Testing venerable ideas against recent theories. *Journal of Public Administration Research and Theory*, 17(3), 357–377.
- Menzel, D.C. (2005). Research on ethics and integrity in governance: A review an assessment. *Public Integrity*, 7(2), 147-168.
- Meynhardt, T. (2009). Public value inside: What is public value creation? *International Journal of Public Administration*, 32(3-4), 192-219.
- Micheli, P., & Neely, A. (2010). Performance measurement in the public sector in England: Searching for the golden thread. *Public Administration Review*, 70(4), 591-600.
- Mishel, L., & Roy, J. (2006). *Rethinking high school graduation rates and trends*. Economic Policy Institute.
- Mitnick, B. (1975). The theory of agency: The policing “paradox” and regulatory behavior. *Public Choice*, 24(1), 27-42.
- Mitchell, M.S., Baer, M.D., Ambrose, M.L., Folger, R., & Palmer, N.F. (2018). Cheating under pressure: A self-protection model of workplace cheating behavior. *Journal of Applied Psychology*, 103(1), 54-73.
- Moore, C., & Gino, F. (2013). Ethically adrift: How others pull our moral compass from true north, and how we can fix it. *Research in Organizational Behavior*, 33, 53-77.
- Moore, M. (1995). *Creating public value: Strategic management in government*. Harvard University Press.
- Mohl, R. (1990). On the edge: Blacks and Hispanics in Metropolitan Miami since 1959. *The Florida Historical Quarterly*, 69(1), 37-56.

- Moynihan, D.P. (2008), *The Dynamics of Performance Management: Constructing Information and Reform*, Georgetown University Press, Washington, DC.
- _____. (2009), "Through a glass, darkly: understanding the effects of performance regimes," *Public Performance & Management Review*, Vol. 32 No. 4, pp. 592-603.
- _____. (2010). A workforce of cynics? The effects of contemporary reforms on public service motivation. *International Public Management Journal*, 13(1), 24-34.
- Moynihan, D., & Beazley, I. (2016). *Toward next-generation performance budgeting: Reflections on the experiences of seven reforming countries*. World Bank. <http://documents.worldbank.org/curated/en/356081478497402740/Toward-next-generation-performance-budgeting-lessons-from-the-experiences-of-seven-reforming-countries>
- Moynihan, D.P., & Landuyt, N. (2009). How do public organizations learn? Bridging cultural and structural perspectives. *Public Administration Review*, 69(6), 1097–1105.
- Moynihan, D. P., Nielsen, P. A., & Kroll, A. (2017). Managerial use of performance data by bureaucrats and politicians. In O. James, S. Jilke, & G. Van Ryzin (Eds.), *Experiments in public management research: Challenges and contributions* (pp. 270–290). Cambridge University Press.
- Moynihan, D.P., & Pandey, S. (2010). The big question for performance management: Why do managers use performance information? *Journal of Public Administration Research and Theory*, 20(4), 849-866.
- Moyson, S., Raaphorst, N., Groeneveld, S., & Van de Walle, S. (2018). Organizational socialization in public administration research: A systematic review and directions for future research. *The American Review of Public Administration*, 48(6), 610-627.
- Mutz, D. C., Pemantle, R., & Pham, P. (2019). The perils of balance testing in experimental design: Messy analyses of clean data. *The American Statistician*, 73(1), 32-42.
- Nagin, D.S., Rebitzer, J.B., Sanders, S., & Taylor, L.J. (2002). Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *The American Economic Review*, 92(4), 850-873.

- Naranjo-Gil, D. (2009). Strategic performance in hospitals: The use of the balanced scorecard by nurse managers. *Health Care Management Review, 34*(2), 161-170.
- Nielsen, P. A., & Baekgaard, M. (2015). Performance information, blame avoidance, and politicians' attitudes to spending and reform: Evidence from an experiment. *Journal of Public Administration Research and Theory, 25*(2), 545–569.
- Nielsen, P. & Moynihan, D. (2017). How do politicians attribute bureaucratic responsibility for performance? Negativity bias and interest group advocacy. *Journal of Public Administration Research and Theory, 27*(2), 269-283.
- Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic Management Journal, 18*(S1), 187–206.
- O'Flynn, J. (2014). From new public management to public value: Paradigmatic change and managerial implications. *Australian Journal of Public Administration, 66*(3), 353-366.
- O'Keefe, D. F., Messervey, D., & Squires, E. C. (2018). Promoting ethical and prosocial behavior: The combined effect of ethical leadership and coworker ethicality. *Ethics & Behavior, 28*(3), 235-260.
- Olsen, A. L. (2015). The numerical psychology of performance information: Implications for citizens, managers, and policymakers. *Public Performance & Management Review 39*(1), 100-115.
- . (2016). Human interest or hard numbers? Experiments on citizens' selection, exposure, and recall of performance information. *Public Administration Review, 77* (3), 408–420.
- Olsen, A. L., Hjorth, F., Harmon, N., & Barfort, S. (2018). Behavioral dishonesty in the public sector. *Journal of Public Administration Research and Theory, 29*(4), 572-590.
- Oorschot, W. V. (2000). Who should get what, and why? On deservingness criteria and the conditionality of solidarity among the public. *Policy & Politics, 28*(1), 33-48.
- Oppel, R. A., & Shear, M. D. (2014, May 28). Severe report finds V.A. hid waiting lists at hospitals. *The New York Times*. <https://www.nytimes.com/2014/05/29/us/va-report-confirms-improper-waiting-lists-at-phoenix-center.html>

- O'Reilly III, C.A., & Chatman, J.A. (1996). Culture as social control: Corporations, cults, and Commitment. In B.M. Staw, & LL. Cummings (Eds.), *Research in Organizational Behavior* (pp. 157-200), JAI Press.
- O'Reilly III, C.A., Chatman, J., & D. F. Caldwell. (1991). People and organisational culture: A profile comparison approach to assessing person-organisation fit. *Academy of Management Journal*, Vol. 34, 487–516.
- O'Toole Jr, L. J., & Meier, K. J. (2010). In defense of bureaucracy. *Public Management Review*, 12(3), 341-361.
- _____. (2011). *Public management: Organization, governance, and performance*. Cambridge University Press.
- Ottaviani, M., & Squintani, F. (2006). Naive audience and communication bias. *International Journal of Game Theory*, 35(1), 129–150.
- Pandey, S. K., & Stazyk, E. C. (2008). Antecedents and correlates of public service motivation. In J.L. Perry and A. Hondeghem (Eds.), *Motivation in public management: The call of public service* (pp. 101-117). Oxford: Oxford University Press.
- Peale, S. (1989), *The diseasing of America*, Lexington Books, Lexington, MA.
- Perkins, W. S., & Rao, R. C. (1990). The role of experience in information use and decision making by marketing managers. *Journal of Marketing Research*, 27(1), 1-10.
- Perrow, C. (1986). *Complex organizations: A critical essay*. Random House
- Perry, J. (1996). Measuring public service motivation: An assessment of construct reliability and validity. *Journal of Public Administration Research and Theory*, 6(1), 5-22.
- Perry, J. L., Hondeghem, A., & Wise, L. R. (2010). Revisiting the motivational bases of public service: Twenty years of research and an agenda for the future. *Public Administration Review*, 70(5), 681-690.
- Perry, J.L., & Wise, L.R. (1990). The motivational bases of public service. *Public Administration Review*, 50, 367–373.
- Radin, B. (2006). *Challenging the performance movement: Accountability, complexity, and democratic values*. Georgetown University Press.

- Radnor, Z. (2008a). Hitting the target and missing the point? Developing an understanding of organizational gaming. In *Performance Information in the Public Sector* (pp. 94-105). Palgrave Macmillan.
- _____. (2008b). Muddled, massaging, manoeuvring or manipulated? A typology of organizational gaming. *International Journal of Productivity and Performance Management*, 57(4), 316-328.
- Rigdon, M.L., & D' Esterre, A.P. (2015). The effects of competition on the nature of cheating behavior. *Southern Economic Journal*, 81(4), 101.
- Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied linguistics*, 22(1), 27-57.
- Rong, X. L., & Grant, L. (1992). Ethnicity, generation, and school attainment of Asians, Hispanics, and non-Hispanic Whites. *Sociological Quarterly*, 33(4), 625-636.
- Rose, J. (2015, April 5). Despite laws and lawsuits, quota-based policing lingers. *National Public Radio*. <https://www.npr.org/2015/04/04/395061810/despite-laws-and-lawsuits-quota-based-policing-lingers>
- Rosen, M. A., Salas, E., Lyons, R., & Fiore, S. M. (2009). Expertise and naturalistic decision making in organizations: Mechanisms of effective decision making. In G. P. Hodgkinson, & W. H. Starbuck (Eds.), *The Oxford Handbook of Organizational Decision Making* (pp. 211-230). Oxford University Press.
- Ross, S. (1973). The economic theory of agency: The principal's problem. Papers and proceedings of the Eighty-Fifth Annual Meeting of the American Economic Association. *The American Economic Review*, 63(2), 134-139.
- Rubin, M., & Hewstone, M. (1998). Social identity theory's self-esteem hypothesis: A review and some suggestions for clarification. *Personality and Social Psychology Review*, 1(2), 40-62.
- Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domain. *Journal of Personality and Social Psychology*, 57(5), 749-761.
- Sabatier, P. (1978). The acquisition and utilization of technical information by administrative agencies. *Administrative Science Quarterly*, 23(3), 396-417.
- Salas, E., Rosen, M. A., & DiazGranados, D. (2010). Expertise-based intuition and decision making in organizations. *Journal of Management*, 36(4), 941-973.

- Schafft, K.A., & Biddle, C. (2013). Place and purpose in public education: School district mission statements and educational (dis) embeddedness. *American Journal of Education*, 120(1), 55–76.
- Schindler, S., & Pfattheicher, S. (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology*, 69, 172–177.
- Schocket, P., & Burghardt, J. (2008). Do job corps performance measures track program impacts? *Journal of Policy Analysis and Management*, 27(3), 556-576.
- Schraeder, M., Tears, R.S., & Jordan, M.H. (2005). Organizational culture in public sector organizations: Promoting change through training and leading by example. *Leadership & Organization Development Journal*, 26(6), 492-502.
- Schweitzer, M.E. and Hsee, C.K. (2002), “Stretching the truth: elastic justification and motivated communication of uncertain information”, *Journal of Risk and Uncertainty*, Vol. 25 No. 2, pp. 185-201.
- Schwieren, C., & Weichselbaumer, D. (2010). Does competition enhance performance or cheating? A laboratory experiment. *Journal of Economic Psychology*, 31(3), 241-253.
- Senn, S. J. (1989). Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine*, 8(4), 467-475.
- Shalvi, S. (2012). Dishonestly increasing the likelihood of winning. *Judgment and Decision Making*, 7(3), 292–303.
- Shalvi, S., Dana, J., Handgraaf, M.J., & De Dreu, C.K. (2011). Justified, ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*, 115(2), 181–190.
- Shu, L.L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M.H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences*, 109(38), 15197-15200.
- Siverbo, S., Cäker, M., & Akesson, J. (2019). Conceptualizing dysfunctional consequences of performance measurement in the public sector. *Public Management Review*, 21(12), 1801-1823.
- Smith, M. (1998). Culture and organisational change. *Management Accounting*, 76(7), 60–62.
- Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18(2). 277-310.

- Solocheck, J. (2018, January 17). Four Hillsborough County schools must lose their principals, Florida Board of Education says. *The Tampa Bay Times*.
<http://www.tampabay.com/blogs/gradebook/2018/01/17/four-hillsborough-county-schools-must-lose-their-principals-florida-board-of-education-says/>
- Sonnentag, S., & Grant, A.M. (2012). Doing good at work feels good at home, but not right away: When and why perceived prosocial impact predicts positive affect. *Personnel Psychology*, 65(3), 495-530.
- Soss, J., Fording, R., & Schram, S. F. (2011). The organization of discipline: From performance management to perversity and punishment. *Journal of Public Administration Research and Theory*, 21(2), i203-i32.
- Strauss, V. (2018, February 1). The shocking thing about D.C.'s schools scandal: And why it has national significance." *The Washington Post*.
https://www.washingtonpost.com/news/answersheet/wp/2018/02/01/the-shocking-thing-about-d-c-s-exploding-schools-scandal-and-why-it-has-national-significance/?utm_term=.82ea2f1679da
- Stritch, J. M., Pedersen, M. J., & Taggart, G. (2017). The opportunities and limitations of using Mechanical Turk (Mturk) in public administration and management scholarship. *International Public Management Journal*, 20(3), 489-511.
- Takeuchi, R., Bolino, M. C., & Lin, C. C.. (2015). Too many motives? The interactive effects of multiple motives on organizational citizenship behavior. *Journal of Applied Psychology*, 100(4), 1239-1248.
- Tarakci, M., Ateş, N. Y., Floyd, S. W., Ahn, Y., & Wooldridge, B. (2018). Performance feedback and middle managers' divergent strategic behavior: The roles of social comparisons and organizational identification. *Strategic Management Journal* 39(4), 1139–1162.
- Taylor, E. Z., & Curtis, M. B. (2018). Mentoring: A path to prosocial behavior. *Journal of Business Ethics*, 152(4), 1133-1148.
- Taylor, J. (2014). Organizational culture and the paradox of performance management. *Public Performance & Management Review*, 38(1), 7-22.
- Tenbrunsel, A.E. (1998). Misrepresentation and expectation of misrepresentation in an ethical dilemma, the role of incentives and temptation. *Academy of Management Journal*, 41(3), 330–339.

- Thau, S., Derfler-Rozin, R., Pitesa, M., Mitchell, M.S., & Pillutla, M.M. (2015). Unethical for the sake of the group: Risk of social exclusion and pro-group unethical behavior. *Journal of Applied Psychology*, 100(1), 98.
- Travis, S. (2016, July 12). 79 South Florida elementary schools face longer days to teach reading. *The Sun-Sentinel*. <http://www.sun-sentinel.com/local/broward/fl-lowest-300-schools-20160712-story.html>.
- Utikal, V., & Fischbacher, U. (2013). Disadvantageous lies in individual decisions. *Journal of Economic Behavior & Organization*, 85, 108-111.
- Vanberg, C. (2008). Why do people keep their promise? An experimental test of two explanations. *Econometrica*, 176(6), 1467–1480.
- Van de Walle, S. (2009). International comparisons of public sector performance. *Public Management Review*, 11(1), 39-56.
- Van de Walle, S., & Bovaird, T. (2007). *Making better use of information to drive improvement in local public services: A report for the audit commission*. Institute of Local Government Studies School of Public Policy, University of Birmingham. <https://www.ssrn.com/abstract=1347349>
- Van Dooren, W., De Caluwe, C., & Lonti, Z. (2012). How to measure public administration performance: A conceptual model with applications for budgeting, human resources management, and open government. *Public Performance & Management Review*, Vol. 35(3), 489-508.
- Van Dooren, W., & Van de Walle, S. (2008). *Performance information in the public sector: How it is used*. Palgrave Macmillan.
- Vanhove, J. (2015). Analyzing randomized controlled interventions: Three notes for applied linguists. *Studies in Second Language Learning and Teaching* 1, 135-152.
- Van Thiel, S., & Leeuw, F.L. (2002). The performance paradox in the public sector. *Public Performance & Management Review*, 25(3), 267-281.
- Verbeke, W., Volgering, M., & Hessels, M. (1998). Exploring the conceptual expansion within the field of organizational behavior: organizational climate and organizational culture. *Journal of Management Studies*, 35(3), 303–330.
- Vigoda-Gadot, E., & Beerli, I. (2011). Change-oriented organizational citizenship behavior in public administration: The power of leadership and the cost of organizational politics. *Journal of Public Administration Research and Theory*, 22(3), 573-596.

- Vincent, L.C., Emich, K.J., & Goncalo, J.A. (2013). Stretching the moral gray zone: Positive affect, moral disengagement, and dishonesty. *Psychological Science*, 24(4), 595–599.
- Walker, R. M., Lee, M. J., James, O., & Ho, S. M. (2018). Analyzing the complexity of performance information use: Experiments with stakeholders to disaggregate dimensions of performance, data sources, and data types. *Public Administration Review*, 78(6), 852-863.
- Walsh, J. P. (1988). Selectivity and selective perception: An investigation of managers' belief structures and information processing. *Academy of Management Journal*, 31(4), 873-896.
- Webeck, S., & Nicholson-Crotty, S. (2019). How historical and social comparisons influence interpretations of performance information. *International Public Management Journal*, 1-24.
- Weibel, A., Rost, K. & Osterloh, M. (2010). Pay for performance in the public sector benefits and (hidden) costs. *Journal of Public Administration Research and Theory*, 20 (2), 387-412.
- Weiss, J. A., & Piderit, S. K. (1999). The value of mission statements in public agencies. *Journal of Public Administration Research and Theory*, 9(2), 193–224.
- Welsh, D.T., & Ordóñez, L.D. (2014). Conscience without cognition: The effects of subconscious priming on ethical behavior. *Academy of Management Journal*, 57(3), 723-742.
- Wilson, J. Q. (1989). *Bureaucracy: What government agencies do and why they do it*. Basic Books.
- Wiltermuth, S.S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes*, 115(2), 157-168.
- Windsor, C.A., & Ashkanasy, N.M. (1996). Auditor independence decision making: The role of organizational culture perceptions. *Behavioral Research in Accounting*, 9, 80–97.
- Woolum, J. (2011). Citizen involvement in performance measurement and reporting: A comparative case study from local government. *Public Performance & Management Review*, 35(1), 79-102.

- Wright, B. E., Hassan, S., & Park, J. (2016). Does a public service ethic encourage ethical behavior? Public service motivation, ethical leadership and the willingness to report ethical problems. *Public Administration*, 94(3), 647–663.
- Yang, K. (2009). Examining perceived honest performance reporting by public organizations: bureaucratic politics and organizational practice. *Journal of Public Administration Research and Theory*, 19(1), 81-105.
- Yoon, K., & Kim, C. (2016). Components of expertise of public officials in emergency management in Korea: An exploratory study. *International Review of Public Administration*, 21(1), 37-56.
- Zhang, H., Chen, S., Wang, R., Jiang, J., Xu, Y., & Zhao, H. (2017). How upward moral comparison influences prosocial behavioral intention: Examining the mediating role of guilt and the moderating role of moral identity. *Frontiers in Psychology*, 8, 1554.
- Zhang, X., Bartol, K.M., Smith, K.G., Pfarrer, M.D., & Khanin, D.M. (2008). CEOs on the edge: Earnings manipulation and stock-based incentive misalignment. *The Academy of Management Journal*, 51(2), 241-258.
- Zhong, C.B., Bohns, V.K., & Gino, F. (2010). Good lamps are the best police: Darkness increases dishonesty and self-interested behavior. *Psychological Science*, 21(3), 311-314.

Appendix

Appendix 1: Survey Experiment

Appendix 1.A: Scenario intro

Help your school to get the numbers right!



Today, student test scores are an important measure in education. They are used to determine a school's success and whether it needs to implement support programs for students. Imagine this scenario...

You are a **school site administrator** in your district. You are supposed to determine the most accurate average score for your school. This task allows some discretion because in order to get a fair average for the true performance of a school, only the most representative student scores are usually included.

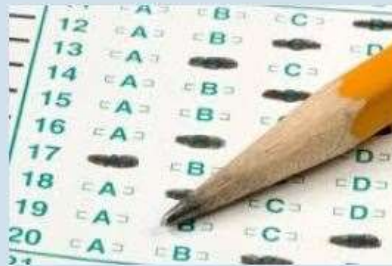
Actual Rewards

Based on a particular task that you will complete in this survey, you or a school in your district may receive actual rewards. You will be informed later in the survey how you can earn the reward.

Click to continue

Appendix 1.A: Scenario intro

What you need to know about test scores



The data you will look at are the scores for reading proficiency. The school district has set its **target for this indicator to 70%**.

70% or more is success for you

If your school meets the target, you will be considered successful as a school administrator.

School average below 70%

If the class average is less than 70%, your school may be considered failing, and your job will be put in jeopardy. When your school does not meet the 70% district target, the school can qualify for the adoption of a reading support program ("SuperStudent"), designed to help students in need achieve.

Click to continue

Appendix 1.B: Treatments

No PI group (control)

Experimental Group 1: NO PI

The SuperStudent program



What is SuperStudent?

SuperStudent is a reading intervention program for students in need of improving their reading skills. It is currently being implemented at your school. SuperStudent has had success at other schools, but because it is a new program for your school, no performance information is available at this time. Click to continue.

The SuperStudent program

Quantitative PI group

SuperStudent Results 2015-2016 School Year

The information below shows how your involvement with SuperStudent affected grade level promotion at your school. It shows what grade level promotion was like before you helped to implement SuperStudent (2013 and 2014) and after you helped to implement the program (2015 and 2016). Information for another school in your school's neighborhood is also available. The other school did not have SuperStudent.

After reviewing the information below, scroll down to see additional information.



Percentage Promoted to the Next Grade*

School	2013	2014	2015	2016
Your School	46%	44%	65%	68%
The Other School	56%	56%	58%	60%

* 2015 & 2016 -SuperStudent implemented at your school but not the other school.

SuperStudent Results 2015-2016 School Year

Here is additional performance information about SuperStudent.

Here you see the percent of students reading on grade level before you helped to implement SuperStudent (2013 and 2014) and after you helped to implement the program (2015 and 2016). Information for the other school is also available. The other school did not have SuperStudent.

After reviewing the information, scroll down to respond to the statements below.



Percentage of Students Reading on Grade Level*

School	2013	2014	2015	2016
Your School	50%	48%	65%	69%
The Other School	50%	52%	58%	57%

* 2015 & 2016 -SuperStudent implemented at your school but not the other school.

Qualitative PI group

The SuperStudent program



Read these two testimonials from students who benefited from your involvement with SuperStudent at your school. After reviewing the information below, scroll down to review two additional testimonials.

Testimonial 1:



"I thank you for recommending SuperStudent. My reading skills have greatly improved."

-Beth Goldstein, 4th grader

Testimonial 2:



"Your excellent job as a citizen advisor in getting SuperStudent has greatly impacted my daughter, who now enjoys reading one hour every day. I also enjoyed the parent workshops."

-Tom Baker, Parent of 2nd grader

The SuperStudent program



Here are two more testimonials from others who have benefited from your involvement with SuperStudent. After reading, scroll down to read one more information about the SuperStudent program.

Testimonial 3:



"We have seen an increase in our property values because of the excellent quality schools like your school. And we know your work as a Citizen Advisor to the school principal has been a major reason."

<https://flu.qualtrics.com/ControlPanel/Ajax.php?action=GetSurveyPrintPreview>

5/16

1/30/2018

Qualtrics Survey Software

-Fred Miller, homeowner in the neighborhood where your school is located

Testimonial 4:



"Our businesses are happy to see high quality schools, which motivates us to expand business in this area."

-Business owner in the neighborhood where your school is located

Library Time in the SuperStudent Program



This is a picture of Charles Parker visiting the library after participating in a SuperStudent session at your school. He used to spend 3 hours per week reading at

<https://flu.qualtrics.com/ControlPanel/Ajax.php?action=GetSurveyPrintPreview>

6/16

1/30/2018

Qualtrics Survey Software

the library. Today, Charlie spends 5 hours per week.



Now, scroll down to respond to the statements below.

The SuperStudent program



Read these two testimonials from students who benefited from your involvement with SuperStudent at your school. After reviewing the information below, scroll down to see two additional testimonials.

Testimonial 1:



"I thank you for recommending SuperStudent. My reading skills have greatly improved."

-Shanaya Watkins 4th grader

Testimonial 2:



"Your excellent job as a citizen advisor in getting SuperStudent has greatly impacted my daughter, who now enjoys reading one hour every day. I also enjoyed the parent workshops."

Denisha Johnson Parent of 2nd grader



Here are two more testimonials from others who benefited from your involvement with SuperStudent. After reading, scroll down to read one more information about the SuperStudent program.

Testimonial 3:



"We have seen an increase in our property values because of the excellent quality at your school. And we know your work as a Citizen Advisor to the school principal has been a major reason."

Kishawn Brown, homeowner in the neighborhood where your school is located

<https://su.qualtrics.com/ControlPanel/Ajax.php?action=GetSurveyPrintPreview>

5/15

1/30/2018

Qualtrics Survey Software

Testimonial 4:



"Our businesses are happy to see high quality schools, which motivates us to expand business in this area."

Kimai Douglas, business owner in the neighborhood where your school is located

Library Time in the SuperStudent Program

This is a picture of Devontay King visiting the library after participating in a SuperStudent session. He used to spend 3 hours per week reading at the library. Today, Devontay spends 5 hours per week.



Appendix 1.C.1: Task – Front end (capturing dependent variable)

Drag and Drop

Time to make your decision. You could keep or transfer some, none, or all students. The individual score for each student is next to the student's name. Drag and Drop each student's name to either: Transfer or Keep.

After you have made your selections, please click to continue

	Transfer Student	Keep Student
Alice Avery 63%		
Bruce Fenders 60%		
Deeana Price 40%		
Shana Burns 33%		
Kim Johnson 17%		
Richard Winter 69%		
Dennis Washington 29%		
Opal Knight 59%		
Bertha Roberts 66%		
Valdimir Kadovsky 62%		

Appendix 1.C.2: Task – Back end (gaming score sheet)

	Transfer Student	Keep Student
Alice Avery 63%	1	0
Bruce Fenders 60%	1	0
Deeana Price 40%	2	0
Shana Burns 33%	2	0
Kim Johnson 17%	3	0
Richard Winter 69%	1	0
Dennis Washington 29%	3	0
Opal Knight 59%	1	0
Bertha Roberts 66%	1	0
Valdimir Kadovsky 62%	1	0

Appendix 2: Supplementary Analysis

The tables in Appendix 2 are supplementary analyses for the experimental study in Essay 3 of this dissertation. Below is an explanation of each table.

Table 10 shows the treatment-demographic correlation. The table shows that the different groups were balanced and one group did not have significantly more of one variable than another group.

Table 10. Appendix 2.A: Treatment-demographics correlations

	Citizen Sample				School Leader Sample			
	No PI	Quant PI	Qual PI (White)	Qual PI (Black)	No PI	Quant PI	Qual PI (White)	Qual PI (Black)
Minority	0.04	-0.01	-0.03	0.01	0.04	-0.05	0.02	-0.01
Female	0.01	0.06	0.01	-0.07 [^]	-0.03	-0.01	0.06	-0.03
Age	-0.03	-0.05	0.05	0.02	0.02	-0.07	0.04	0.02
Liberal	0.02	0.01	-0.07 [^]	0.03	0.02	-0.03	0.04	-0.02

Note: [^] p<0.10

Table 11 shows the effect of prosocial impact treatment across treatment groups without control variables.

Table 11. Appendix 2.B: The effect of prosocial impact (without controls)

	Citizens (1)	Professionals (2)	Citizens (3)	Professionals (4)
Quantitative PI	-0.43 (0.35)	-0.45 (0.58)	-0.43 (0.35)	-0.45 (0.58)
Qualitative PI (combined)	-0.96 (0.02)	-0.09 (0.90)		
Qualitative PI (white students)			-0.76 (0.11)	-0.12 (0.88)
Qualitative PI (black students)			-1.18 (0.01)	-0.07 (0.93)
Constant	7.32 (0.00)	4.55 (0.00)	7.32 (0.00)	4.55 (0.00)
	N	758	286	758
	R ²	0.008	0.001	0.008
				263
				0.001

Note: p-values in parentheses (significant results marked in bold for hypothesized effects); control group serves as reference category; robust standard errors.

Table 12 shows the attention interaction for the novice (general population) sample.

Table 12. Appendix 2.C: Does Attention Matter?

	(1)	(2)
Quantitative PI	-0.054 (0.937)	0.039 (0.954)
Qualitative PI (white students)	-0.657 (0.347)	-0.675 (0.314)
Qualitative PI (black students)	-1.003 (0.714)	-1.036 (0.129)
Attention	0.010 (0.275)	-0.005 (0.954)
Minority (Nonwhite)		0.632 (0.121)
Female		-1.085 (0.002)**
Age		-0.395 (0.019)*
Liberal		-0.188 (0.002)**
North American		-1.273 (0.002)**
Interactions		
Quantitative x Attention	-0.004 (0.711)	-0.005 (0.626)
Qualitative (white) x Attention	0.003 (0.776)	-0.002 (0.830)
Qualitative (black) x Attention	0.001 (0.927)	-0.000 (0.973)
Constant	7.75** (0.00)	10.899** (0.000)
N	758	758
R ²	0.021	0.0

Notes: ** $p < 0.1$, * $p < 0.05$; p-values in parentheses (significant results marked in bold for hypothesized effects); control group serves as reference category; control variables: Minority, Female, Age, Liberal; robust standard errors. The Attention variable was mean-centered to allow more straightforward interpretations of the main effects.

Table 13 shows the effect when administrators were of the same race as students.

Table 13. Appendix 2.D: Student-Administrator Match Based on Race

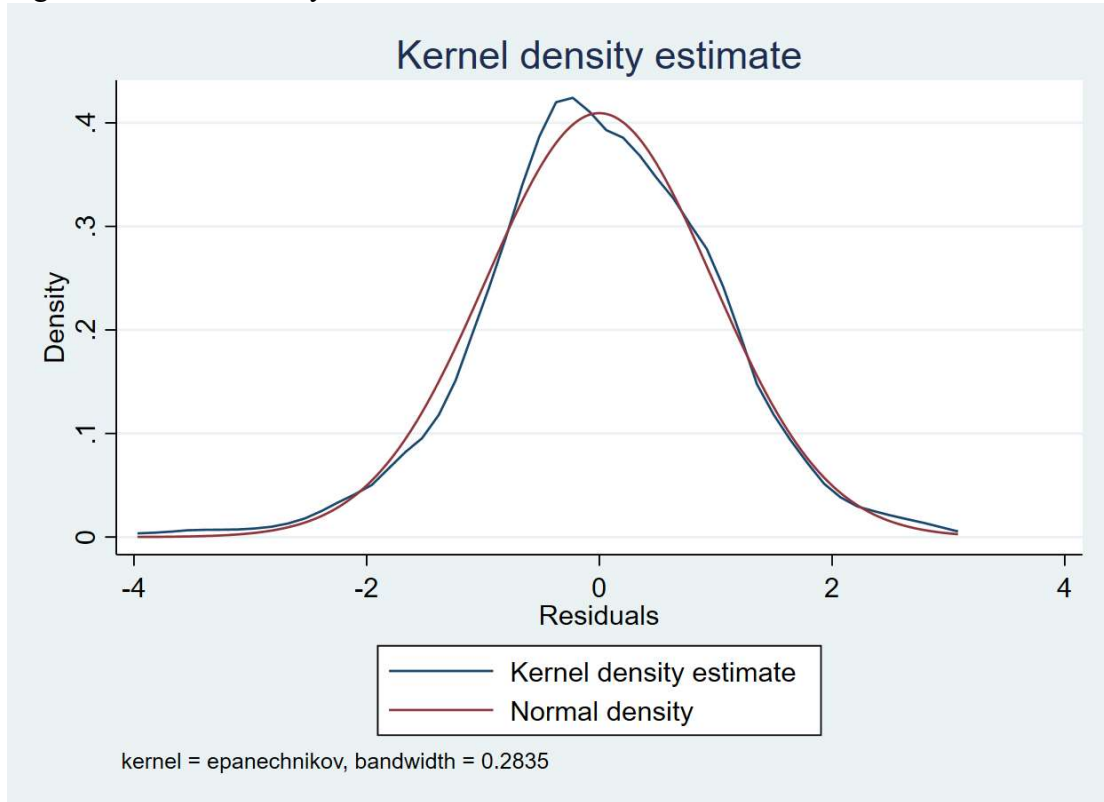
	(1)	(2)
Quantitative PI	-0.790 (0.371)	-0.789 (0.376)
Qualitative PI (white students)	0.215 (0.817)	0.321 (0.736)
Qualitative PI (black students)	0.080 (0.929)	-0.149 (0.868)
Black Admin (incl. Black Hispanic)	0.522 (0.371)	0.719 (0.669)
Female		-0.852 (0.260)
Age		-0.144 (0.684)
Liberal		-0.157 (0.216)
Interactions		
Quantitative x Black Admin.	-0.862 (0.691)	0.670 (0.759)
Qualitative (white) x Black Admin.	-1.546 (0.495)	-1.565 (0.490)
Qualitative (black) x Black Admin.	0.134 (0.952)	0.454 (0.839)
Constant	4.264** (0.00)	6.267** (0.000)
N	272	263
R ²	0.010	0.022

Notes: ** p < 0.1, * p < 0.05; p-values in parentheses (significant results marked in bold for hypothesized effects); control group serves as reference category; control variables: Female, Age, Liberal; robust standard errors.

Appendix 3: OLS Regression Diagnostics for Essay 2 – Effort Substitution

Appendix 3 shows diagnostic tests for the OLS Regression used in Essay 2, the study on effort substitution. The first test shown in Figure 5 is the test for normality of the residuals using the Kernel density test. The estimate shows normality of the residuals.

Figure 4: Kernel Density Plot of Residuals for Effort Substitution



Tables 14 and 15 show the Breusch-Pagan/Cook-Weisberg test for heteroscedasticity. Since $p < 0.05$, I rejected the null hypothesis and conclude there is a possibility of heteroscedasticity, which means there is an issue. To address this issue, I use robust standard errors in the main regression tables. I clustered standard errors for schools. This is explained in the footnote on page 72. Using robust standard errors also addresses the issue of independence of the residuals.

Table 14: Breusch-Pagan / Cook-Weisberg Test for Heteroscedasticity Grad Gaming

Graduation Gaming	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
Percent Low Income	-0.0046307	0.0033616	-1.38	0.17	-0.0112502	0.0019889
Percent Black	0.0069763	0.0025912	2.69	0.008	0.0018738	0.0120788
Charter School	0.0482808	0.1973599	0.24	0.807	-0.3403536	0.4369151
Performance Grade	0.0652211	0.0650528	1	0.317	-0.0628786	0.1933207
Size	-0.0000689	0.0000758	-0.91	0.364	-0.0002181	0.0000803
Climate	-0.0291445	0.1502758	-0.19	0.846	-0.3250624	0.2667734
Performance Use	-0.1580293	0.189	-0.84	0.404	-0.5302015	0.214143
Lack of Resources	-0.1352805	0.1111826	-1.22	0.225	-0.3542174	0.0836565
_cons	1.211822	1.032491	1.17	0.242	-0.8213242	3.244969

Breusch-Pagan / Cook-Weisberg test for heteroscedasticity

Ho: Constant variance

Variables: fitted values of gradgame

chi2(1) = 4.94

Prob > chi2 = 0.0263

Table 15: Breusch-Pagan/Cook-Weisberg Test for Heteroskedasticity Acc. Gaming

Acceleration Gaming	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
Percent Low Income	0.0054554	0.0047665	1.14	0.253	-0.0039303	0.0148412
Percent Black	0.01058	0.0036827	2.87	0.004	0.0033283	0.0178317
Charter School	-1.317184	0.2800452	-4.7	0	-1.868629	-0.7657387
Performance Grade	0.1176754	0.0924622	1.27	0.204	-0.0643946	0.2997454
Size	-0.0002156	0.0001076	-2	0.046	-0.0004276	-3.70E-06
Climate	0.0987069	0.2136266	0.46	0.644	-0.3219517	0.5193655
Performance Use	-0.2072595	0.2686864	-0.77	0.441	-0.736338	0.3218191
Lack of Resources	-0.150189	0.1578144	-0.95	0.342	-0.4609461	0.1605681
_cons	0.5973752	1.467568	0.41	0.684	-2.292456	3.487207

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: fitted values of gradgame

chi2(1) = 4.94

Prob > chi2 = 0.0263

Table 16 through Table 19 show robust regression for the two gaming measures (graduation gaming and acceleration gaming). For each gaming measure, there are two tables (one for each mission measure: charter and mission specificity). The robust regression tables, whose estimator is not based on the OLS assumptions, show that the findings are about the same as with the OLS regression.

Table 16: Graduation Gaming Model 1: Charter School

Robust Regression					Number of obs = 268	
					F(12, 255) = 6.54	
					Prob > F = 0.00	
Graduation Gaming	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
Percent Low Income	-0.0112415	0.002029	-5.54	0.000	-0.0152373	-0.00725
Percent Black	0.0061272	0.00147	4.17	0.000	0.0032328	0.009022
Charter School	-0.1634876	0.102819	-1.59	0.113	-0.3659692	0.038994
Performance Grade	0.0010217	0.042371	0.02	0.981	-0.0824207	0.084464
Size	-0.0001272	3.86E-05	-3.30	0.001	-0.0002032	-5.1E-05
Climate	-0.0164428	0.089063	-0.18	0.854	-0.191836	0.15895
Performance Use	-0.1670555	0.108199	-1.54	0.124	-0.3801323	0.046021
Lack of Resources	-0.0187620	0.066004	-0.28	0.776	-0.1487436	0.11122
Year						
2013	0.1244535	0.115281	1.08	0.281	-0.1025706	0.351478
2014	-0.0163146	0.111484	-0.15	0.884	-0.2358605	0.203231
2015	0.1020392	0.112302	0.91	0.364	-0.1191185	0.323197
2016	0.0209867	0.11854	0.18	0.860	-0.2124547	0.254428
_cons	1.5781190	0.633516	2.49	0.013	0.33052900	2.825709

Table 17: Graduation Gaming Model 2: Mission Specificity

Robust Regression					Number of obs = 268	
					F(12, 255) = 6.54	
					Prob > F = 0.00	
Graduation Gaming	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
Percent Low Income	-0.0112415	0.002029	-5.54	0.000	-0.015237	-0.00725
Percent Black	0.0061272	0.001470	4.17	0.000	0.003233	0.00902
Mission Specificity	-0.1634876	0.102819	-1.59	0.113	-0.365969	0.03899
Performance Grade	0.0010217	0.042371	0.02	0.981	-0.082421	0.08446
Size	-0.0001272	0.000039	-3.3	0.001	-0.000203	-0.00005
Climate	-0.0164428	0.089063	-0.18	0.854	-0.191836	0.15895
Performance Use	-0.1670555	0.108199	-1.54	0.124	-0.380132	0.04602
Lack of Resources	-0.0187620	0.066004	-0.28	0.776	-0.148744	0.11122
Year						
2013	0.1244535	0.115281	1.08	0.281	-0.1025706	0.3514776
2014	-0.0163146	0.1114836	-0.15	0.884	-0.2358605	0.2032313
2015	0.1020392	0.1123021	0.91	0.364	-0.1191185	0.3231970
2016	0.0209867	0.1185397	0.18	0.860	-0.2124547	0.2544281
_cons	1.5781190	0.633516	2.49	0.013	0.33052900	2.8257090

Table 18: Acceleration Gaming Model 1: Charter School

Robust Regression					Number of obs = 268	
					F(12, 255) = 19.72	
					Prob > F = 0.00	
Graduation Gaming	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
Percent Low Income	0.009605	0.0024851	3.87	0.00	0.004711	0.014499
Percent Black	0.005110	0.0018765	2.72	0.01	0.001414	0.008805
Charter School	-1.467240	0.143559	-10.22	0.00	1.749952	-1.18453
Performance Grade	0.084767	0.0514751	1.65	0.10	0.016604	0.186137
Size	-0.000263	0.0000552	-4.76	0.00	0.000371	-0.00015
Climate	0.254891	0.1096224	2.33	0.02	0.03901	0.470771
Performance Use	-0.217056	0.1377807	-1.58	0.12	0.488389	0.054277
Lack of Resources	-0.007897	0.0811153	-0.1	0.92	0.167638	0.151845
Year						
2013	0.207797	0.1410714	1.47	0.14	-0.07002	0.485611
2014	0.143415	0.1363051	1.05	0.29	0.125012	0.411842
2015	0.265177	0.1361824	1.95	0.05	0.003008	0.533363
2016	0.262470	0.1439798	1.82	0.07	0.021071	0.546011
_cons	-0.593210	0.7861901	-0.75	0.45	2.141463	0.9550424

Table 19: Acceleration Gaming Model 2: Mission Specificity

Robust Regression					Number of obs = 268	
					F(12, 255) =	7.91
					Prob > F =	0.00
Acceleration Gaming	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
Percent Low Income	0.0075532	0.0029605	2.55	0.011	0.0017231	0.0133834
Percent Black	0.0097233	0.0021445	4.53	0.000	0.0055002	0.0139464
Mission Specificity	-0.437538	0.1500164	-2.92	0.004	-0.7329668	-0.1421092
Performance Grade	0.0618169	0.0618215	1.00	0.318	-0.0599289	0.1835627
Size	-7.91E-06	0.0000563	-0.14	0.888	-0.0001188	0.000103
Climate	0.1309105	0.1299469	1.01	0.315	-0.1249953	0.3868162
Performance Use	-0.5930932	0.1578663	-3.76	0.000	-0.9039809	-0.2822054
Lack of Resources	-0.1071792	0.0963019	-1.11	0.267	-0.2968276	0.0824691
Year						
2013	0.2400387	0.1681996	1.43	0.155	-0.0911985	0.571276
2014	0.1188672	0.1626591	0.73	0.466	-0.2014591	0.4391935
2015	0.2064285	0.1638533	1.26	0.209	-0.1162496	0.5291065
2016	0.1734438	0.1729541	1.00	0.317	-0.1671566	0.5140442
_cons	0.9244459	0.9243253	1.00	0.318	-0.8958377	2.74473

VITA

SHAWN LORENZO BENAINE

Born, Kingston, Jamaica

2000-2004

B.S., Mass Communications
Florida International University
Miami, Florida

2005-2006

M.S., Public Affairs
Florida International University
Miami, Florida

2006-2015

K-12 Teacher
Miami-Dade County Public Schools
Miami, Florida

2015-2020

Teaching & Research Assistant
Florida International University
Miami, Florida

2015-2020

Doctoral Candidate
Florida International University
Miami, Florida

2020

Research & Evaluation Manager
Children's Services Council of Broward County
Lauderhill, Florida

PUBLICATIONS AND PRESENTATIONS

Benaine, S. Lorenzo and Alexander Kroll. (2019). *Explaining effort substitution in performance systems: The role of task demands and mission orientation*. Public Management Review, 1-23

Benaine, S. Lorenzo (2020). *Performance gaming: A systematic review of the literature in public administration and other disciplines with directions for future research*, International Journal of Public Sector Management.

Benaine, S. Lorenzo and Alexander Kroll. *Does prosocial impact reduce data gaming?* Presented at 2019 Public Management Research Conference (PMRC), UNC-Chapel Hill, NC.