2020

How Historical and Social Comparisons Influence Interpretations of Performance Information

Webeck, Sean; Nicholson-Crotty, Sean

Taylor & Francis

http://hdl.handle.net/10945/68611

International Public Management Journal

HOW HISTORICAL AND SOCIAL COMPARISONS INFLUENCE INTERPRETATIONS OF PERFORMANCE INFORMATION

SEAN WEBECK
ANAVAL POSTGRADUATE SCHOOL

SEAN NICHOLSON-CROTTY
BINDIANA UNIVERSITY

ABSTRACT: *The ways in which managers use performance information is among the most salient topics in the study of public management. Drawing inspiration from several recent studies on the use of performance information by citizens, we adopt a behavioral approach to understand the influence of reference dependence on the interpretation of performance information by managers. Specifically, we run two experiments in a sample of professional managers, which allows us to test whether social and historical comparisons influence how respondents interpret performance information. The results suggest that framing an objective performance metric as poor relative to peer or competitor organizations leads managers to rate overall organizational performance significantly lower than managers in a control group who received the same metric, but no comparative frame. The results support expectations about the importance of social comparisons, particularly in the case of negative deviations from the reference point. The fact that we find no impact of historical comparisons on the interpretation of performance information deviates from recent work on citizen evaluations and suggests differences in the ways in which elites process such information. We conclude with a discussion of the implications of our results for the study of performance management and the behavioral approach to public management more generally.*

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/upmj.

International Public Management Journal, Vol 23(6), pages 798–822
DOI: 10.1080/10967494.2018.1550129

ISSN: 1096-7494 print/1559-3169 online

INTRODUCTION

In 1995, Robert Behn suggested that one of the big questions facing scholars who study public sector organizations was understanding how "*public managers* use measures of the achievements of public agencies to produce even greater achievements" (Behn 1995, emphasis added). Yet recently, 20 years after Behn pointed out a major question for researchers in the field, a prominent scholar commented that "we know little about the basic tendency of individuals to incorporate and use performance information" (Moynihan 2015, 33). Kroll (2015a, 2015b) made a significant contribution to this literature with his review of how managers use performance information, but noted that very few studies incorporated a psychological perspective to understand how public managers *process* performance information (note, as exceptions, Kroll 2015a, 2015b; Andersen and Moynihan 2016).

We seek to contribute to this literature by focusing on the question of relative performance evaluation. Research suggests that people do not judge information about performance in absolute terms, but in reference to how an organization performed previously (historical comparison) or how it performed relative to peer institutions (social comparison). To date, however, scholars have focused primarily on the historical and social comparisons made by ordinary citizens and not by organizational managers. Because we know that elites process information differently from average citizens due to expertise and experience, among other factors, these results may not offer an accurate picture of the ways in which public managers use these heuristics when judging performance information.

To address this potential gap in our understanding we adopt a behavioral approach to understand the influence of reference dependence on the interpretation of performance information. Specifically, we run two experiments in a sample of professional managers, which allow us to test whether social and historical comparisons influence how respondents interpret performance information. In line with previous research on citizens (Charbonneau and Van Ryzin 2015; Olsen 2017), we provide evidence that social comparisons do more to shape how professional managers evaluate and interpret performance information than historical comparisons. Distinct from studies of citizens, however, we find no evidence that historical comparisons have a significant influence on the way in which professional managers interpret performance information. These results provide potentially important insights regarding the design of systems to provide performance information to managers and suggest ways in which performance information should be framed for different audiences. Moreover, the discrepancies between our results and those of recent research suggest that findings from studies of citizens may not tell us everything we need to know about the use of performance information by managers.

The remainder of the article proceeds in four parts. First, we review the scholarly literature on the use of performance information. Next, we discuss the advantages of an experimental approach and explain our experimental research design. We then describe two experiments and present their results. Finally, we conclude with a discussion and some practical and scholarly implications of this research.

PERFORMANCE INFORMATION

Bureaucrats and the "Use" of Performance Information

From both academic and practical perspectives, one of the compelling aspects of performance management systems is that they generate, at least ostensibly, an objective assessment of how well an organization is doing its job (Nielsen 2013). And yet, *who* or *which organization* assesses performance metrics potentially influences how the information is interpreted (Moynihan 2008). The subjectivity of performance information use is one of the key reasons why significant questions remain regarding the ways in which "objective" performance metrics influence organizational outcomes.

While scholars study the use of performance information across three groups of "end users"—citizens, managers, and politicians (Van de Walle and Van Dooren 2011)—public managers might represent the main or primary users of performance information. At the very least, they are a key target of the information that nearly ubiquitous performance measurement systems produce. Despite their importance, there is still much to learn about how public managers use performance information (Kroll 2014, 2015a, 2015b), which is why we make them the focus of this study.

The literature to date on this subject has clearly established the subjectivity of performance information use by government officials. In an early study, Behn (2003) argued that bureaucrats use performance metrics to evaluate, control, budget, motivate, promote, celebrate, learn, and improve. Yet, he emphasized that the last of these—improve—was the most important. In fact, he argued that all the rest "are simply means for achieving this ultimate purpose," which pertains to improving performance (Behn 2003, 588).

Rather than looking at specific actions, Moynihan (2010) suggests that there are four strategies public managers can employ when using performance information: passive, political, perverse, and purposeful. In other work, Moynihan (2008) suggests that performance information is selected and presented to persuade others. For Moynihan, performance information becomes subjective because actors add their own interpretation to the data. He argues that this plays an important role in how we should understand performance information use. As we discuss later, we extend this logic of the importance of interpretation as a part of a cognitive process of performance information use.

A recent systematic literature review offers another set of factors that help us understand why, given their discretion in doing so, bureaucrats sometimes use performance information (Kroll 2015a, 2015b). Over the last 15 years in "a highly relevant and fast-growing research area" (Ibid.:460), research consistently shows that six factors commonly drive the use of performance information among bureaucrats: measurement system maturity (for examples, see Berman and Wang 2000; Ho 2006; Taylor 2009), stakeholder involvement (for examples, see Ho 2006; Bourdeaux and Chikoto 2008; Moynihan and Pandey 2010; Moynihan and Lavertu 2012), leadership support (for examples, see Moynihan and Ingraham

2004; Yang and Hsieh 2007; Moynihan and Lavertu 2012), support capacity (for examples, see Berman and Wang 2000; Julnes and Holzer 2001; Moynihan and Hawes 2012), innovative culture (for examples, see Moynihan 2005; Moynihan and Pandey 2010; Moynihan et al. 2012b), and goal clarity (for examples, see Moynihan and Landuyt 2009; Moynihan et al. 2012a; Moynihan et al. 2012b).

For our purposes, the most important takeaway from this review of previous work on performance information use is that many studies of how bureaucrats use performance information look to organizational behavior, organizational theory, or a combination of these approaches to explain the phenomenon. In other words, they focus mainly on the ways in which the organizational context of performance measurement systems influences the use of performance information.

Alternatively, recent scholarship suggests that *individual behavioral* factors might produce systematic variation in the use of performance information among bureaucrats (Kroll 2015a, 2015b; Moynihan et al. 2017). For example, several studies over the past decade suggest that psychology may play a role in helping us understand how bureaucrats use performance information (Moynihan 2008; Salge 2011; Nielsen 2013; Kroll 2015a, 2015b; Moynihan 2015; Andersen and Moynihan 2016). In addition to these, some very recent pieces also demonstrate the utility of incorporating an individual-level behavioral approach to examine the use of performance information.

A significant majority of these studies look at how citizens respond to performance information (Andersen and Hjortskov 2016; Baekgaard and Serritzlew 2016; Barrows et al. 2016; Hvidman and Andersen 2016; Olsen 2013, 2015a, 2017), but other work has looked at how politicians (Olsen 2014; Nielsen and Baekgaard 2015; George et al. 2016; Nielsen and Moynihan 2017) use performance information as well. Importantly, some of this work has taken place in the context of education, which is the same service area we focus on in our experiments. These studies have demonstrated that benchmarking plays an important role in shaping the cognitive processing of performance information.

There are three important takeaways from these studies. First, experimental methods are a useful approach to develop our understanding of performance information use across a variety of political actors (Anderson and Edwards 2015; Bouwman and Grimmelikhuijsen 2016; Jilke et al. 2016; James et al. 2017). Second, when it comes to performance metrics, these studies suggest that, depending upon the circumstances, individuals exhibit various cognitive biases and utilize several heuristics when responding to performance information. Evidence for cognitive bias in the use of performance information by various actors supports the value of taking an information processing approach. Finally, an examination of the recent literature reveals that a relatively limited amount of work has taken a behavioral approach to understanding the use of performance information by public managers. Because these people are a primary target of this information, understanding how cognitive biases influence their assessment of performance metrics is, we believe, an important contribution to the literature.

REFERENCE POINTS AND INFORMATION PROCESSING

While there are several cognitive biases that may influence assessments of performance information, we focus on the use of reference points and particularly negative deviations from accepted referents in this study. The idea of reference points in the interpretation of information is nothing new, as we demonstrate in the following, but to date this concept has not been used to understand how public managers interpret performance information.

Psychologists have long understood that human judgment is fundamentally comparative in nature (Mussweiler 2003). More specifically, we know that individuals make temporal (Albert 1977) and social (Festinger 1954) comparisons when evaluating abilities, information, and opinions. Reference points influence decisions because human perception is "*reference-dependent*" (Kahneman 2002:459, emphasis in original). These references serve as "stimuli of known attributes that act as standards against which other categorically similar stimuli of unknown attributes are compared in order to gain information" (Yockey and Kruml 2009:97). Consistent with the idea that limitations in human processing constrain our ability to accept, hold, and process information (Freeman 1954; Simon 1955), reference points also serve as cognitive heuristics in making evaluative judgments about information (Mussweiler and Epstude 2009; Mussweiler and Posten 2012).

Historical and Social Reference Points, Negativity Bias, and the Interpretation of Performance Information by Professional Managers

For a variety of reasons, it is reasonable to assume that the use of reference points also influences the interpretation of performance information. Indeed, Herbert Simon argued that "the only sound basis for decisions about numbers is numerical factual information about past experiences or the experiences of others—nothing more nor less than comparative statistics" (Simon 1939, 106, as cited in Olsen 2015b, 108). Reference points facilitate those comparisons, but there are numerous comparative reference points that public managers might use as references. Consistent with recent work on citizen evaluations of public sector performance, we focus on historical and social performance comparisons as points of reference in this article (Olsen 2015a). Historical reference points allow comparison of the performance of an organization to the previous performance of the same organization. In other words, past performance provides a salient status quo against which individuals can easily assess change and decide if current performance is acceptable. This type of historical reference point is common in performance measurement systems, such as No Child Left Behind, which individuals can use to judge schools on progress relative to the previous year. Yet, it is important to reiterate that our primary interest is in the degree to which such historical comparisons influence the assessment of performance information. Following studies of citizen interpretation of performance information, we hypothesize that providing information about better (worse) past performance of a public-sector organization

will lower (raise) a bureaucrat's assessment of the organization's current performance. Olsen (2017),

Social reference points provide another frame in which individuals can compare performance. With a social reference point, individuals compare the performance of their organization against the performance of other, comparable organizations at the same point in time. Often, we might think of these as peer organizations, competitors, or simply organizations in a similar geographic region. The power of social comparisons has received a great deal of empirical support. Festinger (1954) suggested that "people evaluate their opinions and abilities by comparison respectively with the opinions and abilities of others" (118). Charbonneau and Van Ryzin (2015) and Olsen (2017) demonstrate that social comparisons can influence the ratings an individual gives to a public organization. We expect that providing information about better (worse) performance relative to other, comparable organizations will lower (raise) a manager's assessment of an organization's current performance.

There is also preliminary evidence to suggest that social comparisons weigh heavily on the thinking of individuals when assessing performance. Evidence to support this argument comes in the form of survey responses of public managers in North Carolina (Ammons and Rivenbark 2008). There is also experimental evidence that citizens potentially give more weight to social rather than historical comparisons (Charbonneau and Van Ryzin 2015; Olsen 2017). We expect that managers will give greater weight to social comparisons than to historical comparisons.

In addition to our expectations about how public managers will interpret performance information when it is presented as a historical or social comparison, we are interested in whether the *direction* of the deviation from the reference point matters. Scholars have long recognized that human beings tend to respond more strongly to negative information than to comparable information that is framed in a positive way (Ito et al. 1998). This is because negative information is generally more salient and more potent than positive information (Rozin and Royzman 2001), thus drawing more cognitive processing.

These expectations have received considerable support in recent research on both citizens' and politicians use of performance information (Boyne et al. 2009; Charbonneau and Bellavance 2012; Hood 2007; James 2011; James and John 2007; James and Moseley 2014; Olsen 2015a; (Nielsen and Baekgaard 2015; Nielsen and Moynihan 2017). We expect that professional managers will also exhibit a negativity bias and respond more strongly to lower performance relative to an accepted reference point than to an equivalent but positive deviation.

EXPERIMENTAL DESIGN AND METHOD

Participants

Data for the two experiments come from a Qualtrics panel collected during May 2017. We recruited respondents directly through Qualtrics to avoid some of the potential pitfalls of using other online survey platforms (Stritch, Mogens, and Taggart

2017). Qualtrics screened and provided the respondents for the survey with the stipulation that respondents were managers in their organization. A partner of Qualtrics initially targeted each of the respondents through self-reporting. Subjects were then screened to remove misidentified respondents using red herrings and other techniques to ensure sample accuracy. Subjects were screened one last time regarding sector, experience, and responsibilities at the beginning of the survey to remove individuals whose answers did not match responses from previous screenings.[1]

The total sample size is 300, with 150 coming from the private sector and 150 from the public sector. On average, our sample is 46 years of age, with more than 25 years in the workforce and more than 10 years in their current positions. The median respondent manages between 100 and 249 people. Almost one-third of members of the sample are responsible for more than 1,000 employees. In other words, these are experienced managers and that is the group to which we can most safely draw inference from the results of this study. Approximately 64% of the sample have at least a bachelor's degree, and 29% completed post-graduate work. The sample is roughly divided between men and women. All subjects work in organizations based in the United States.

Our sample includes managers from both sectors because another experiment in the survey required this sector breakdown. But here, we will primarily focus on the public-sector managers. We pre-registered the survey with the Evidence in Governance and Politics (EGAP) group under the following ID: 20170501AC.

Olsen (2017) ran a set of experiments like ours involving Danish citizens. He had 3,443 respondents for both of his surveys. Our focus on managers allows us to extend this research using, as respondents, a group of individuals who are likely more accustomed than regular citizens to seeing, thinking about, and using performance information in their decision making. Our sample consists of 300 respondents, evenly divided between public and private managers. In Table 1, we provide the descriptive statistics for the combined sample.

Design of the Experiments

In the design of our experiments, we chose to use the substantive policy area of education—specifically, passing rates for standardized test scores—and modeled the performance information on real test score data to make the experiment more plausible and generalizable to real-world decision making. Specifically, we utilized publicly available data from public schools in the state of Indiana.[2] Data for 2011 and 2012 suggested an average change in the pass rate for English and Math exams to be roughly 2%. We averaged the pass rates for schools in the state for both English and Math standardized exams. Since our experiments include historical comparisons, we also compared school test results across time. Doing this led us to our chosen performance metrics—77% passed the English exam and 79% passed the Math exam—as well as the historical comparison data (2% change from last year).

Due to certain constraints of our research project—namely, recruiting a sample of 300 professional managers—we wanted to address two concerns. The first concern involved the potential for the first experiment to influence the way

TABLE 1
Descriptive Statistics

Descriptive Statistics					
Variable	Obs	Mean	Std. Dev.	Min	Max
Full Sample					
Age	300	45.94	12.37	19	74
Education	300	5.16	2.11	1	9
Female	297	0.49	0.50	0	1
Public-Sector Manager	300	0.50	0.50	0	1
Years in Private Sector	217	14.87	12.59	0.5	51
Years in Public Sector	186	17.26	11.41	1	45
Years in Workforce	300	25.25	12.52	1	58
Private-Sector Managers					
Age	150	45.23	121.84	19	74
Education	150	4.76	2.01	1	9
Female	149	0.51	0.50	0	1
Years in Private Sector	126	19.03	13.73	1	51
Years in Public Sector	43	12.52	11.37	1	40
Years in Workforce	150	24.48	12.96	1	58
Public-Sector Managers					
Age	150	46.66	11.89	19	74
Education	150	5.57	2.14	1	9
Female	148	0.47	0.50	0	1
Years in Private Sector	91	9.12	7.81	0.5	37.75
Years in Public Sector	143	18.68	11.08	1	45
Years in Workforce	150	26.02	12.06	1	55

respondents approached the second experiment. We addressed this by separating the two experiments. Specifically, in this regard, respondents saw *Experiment I* near the beginning of the survey and *Experiment II* near the end of the survey. The average response time for the survey across all 300 respondents was 24 minutes and 22 seconds. This meant that there was a significant amount of time between these two experiments. We believe that this was an adequate way to address any concerns about the first experiment influencing the results of the second. Another notable difference between the two experiments was that, in providing the social comparisons in *Experiment II,* we only indicated if the school was in the top or bottom half of local schools. That is, we did not include a rank (e.g., third out of 10). This omission was deliberate and an attempt to help respondents not conflate the two experiments, despite their similarities. It also allowed us to address a potential concern about causal inference. Specifically, if a respondent sees a prompt that reads "top half," we wouldn't know if she interprets this as first out of three or forty-ninth out of 100.

The focus on the educational context in our experiments builds on recent work that takes a behavioral approach to *citizen's* assessments of public organizations. Specifically, it builds on a set of studies that explore the influence of performance targets and relative performance information on those assessments (Barrows et al. 2016; Charbonneau and Van Ryzin 2015).[3] These similarities will allow us to compare the use of performance information by public managers with those of citizens in a comparable service delivery area, which we believe constitutes another contribution for the study.

Before moving on, it is important to note that while we do focus on managers rather than citizens as subjects, our design does not allow us to test the influence of reference effects on their assessments of performance information in their own organizations. Despite this, we believe that the use of managers offers unique insights into the role of historical and social comparisons in performance information use for several reasons. First, whether being asked to consider their own organization or not, professional managers are likely far more familiar with the types and uses of performance information than are citizens. Borrowing from work on political psychology, we can therefore consider professional managers as "sophisticates" because of their experience with performance information. Research suggests that sophisticates process information differently and more effectively than non-sophisticates (Gaines et al. 2007). They make different decisions (Luskin 1987; Mintz, Redd, and Vedlitz 2006) and are better able to connect new information to existing knowledge and to relevant decisions (Jerit, Barabas, and Bolsen 2006). Given these differences, it is reasonable to expect that cognitive biases may influence the assessments of performance information by managers, who are more sophisticated in the use of such information, in systematically different ways than they influence the assessments of citizens.

Experiment I

In Experiment I, we asked respondents to rate the performance of an unnamed high school (High School A) using performance data from a standardized English exam. The goal of Experiment I was to observe the assessed performance when both historical and social comparison information were presented together. We felt that this would be a suitable way to design the experiment for two reasons. First, in a realistic organizational decision-making environment (i.e., a non-experimental setting), managers might have a sense of their organization's performance as well as the performance of peer and competitor organizations. Second, by including both comparison types in the same experimental frame, we might be able to get some sense of the strength of the positive and negative versions of each comparison. Of course, we were also able to compare performance assessments against the control group as well.

In this experiment, individuals saw a raw performance metric which stated that 77% of students at High School A passed the English exam. Respondents were then randomly assigned to one of five groups. The control group saw only the raw performance metric. The other groups saw four combinations of historical and social comparisons. The historical comparison prompts said that the performance was indicative of a 2% increase or decrease in the rate of students who passed the

standardized English exam. The social comparison indicated that, based upon the pass rate, the school ranked third or seventh out of 10 comparable local schools. For the social comparison prompt, individuals were told if this was in the top or bottom half of local schools, respectively. Individuals were then asked to rate the performance of the school using a 0–100 sliding scale.

As an example, someone in the group that saw prompts indicating increases for both the historical and social comparisons saw the following prompt:

> *English Exam: 77% of students in "High School A" passed their standardized English exam. This represents a 2% increase from the previous year. It also means the school was in the top half of local schools in the area (3rd out of 10). Assuming this is the only information available to you, use the sliding scale (0-100) to assess the overall performance of HIGH SCHOOL A over the last year.*

Respondents would rate the performance of the school with the sliding scale. Experimental vignettes for both experiments can be found in the Appendix.

Experiment II

In Experiment II, we asked respondents to rate the performance of an unnamed high school (High School B) using performance data from a standardized Math exam. In this experiment, we wanted to look at the comparisons individually so that we could get a sense of the strength of the comparisons by themselves in the assessment of performance data. Again, respondents were randomly assigned to one of five groups.

To create some generalizability across the two experiments, we used similar comparisons from the first experiment. Individuals saw a raw performance metric that stated that 79% of students at this high school passed the Math exam. As before, the control group saw only the raw performance metric. The other groups saw one of four possible historical and social comparisons. That is, groups 2–5 only saw one of the following: 2% increase from last year, 2% decrease from last year, top half of comparable local schools, or bottom half of comparable local schools. Again, respondents were asked to rate the performance of the school on a 101-point sliding scale.

RESULTS

Our primary analytic strategy includes ANOVA tests, mean difference tables, figures with means and 95% confidence intervals, and Bonferroni cross-group comparison tests, which is consistent with the approach recommended in recent behavioral public administration work (James et al. 2017). As a robustness check, we also include regression models including pretreatment covariates in the Appendix. As Table 2 suggests, however, randomization of subjects across treatment and control groups was adequate. The Qualtrics sample of respondents had both private-

TABLE 2
Randomization Checks

Randomization Checks												
Experiment 1: English Exam: Randomization Check, by Treatment Group												
Treatment	Control		(H) Inc (S) Top		(H) Inc (S) Bot		(H) Dec (S) Top		(H) Dec (S) Bot		ANOVA	
Controls	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	F-statistic	Prob > F
Age	45.12	13.44	45.98	12.11	44.74	12.35	47.18	UL.74	46.79	12.42	0.44	0.78
Education	4.88	1.97	4.96	2.30	5.32	2.11	5.37	2.17	5.25	2.04	0.65	0.63
Female	0.46	0.50	0.46	0.50	0.57	0.50	0.44	0.50	0.51	0.50	0.72	0.58
Yrs for profit	16.61	14.90	15.60	13.40	14.02	11.10	13.74	12.91	14.48	10.55	0.40	0.81
Yrs public	16.40	12.95	19.32	UL11	13.78	10.45	20.15	UL13	17.37	UL02	Z01	0.10
Yrs workforce	24.34	13.35	25.50	12.04	24.03	11.75	26.69	13.09	25.75	12.55	0.46	0.76
Experiment II: Math Exam: Randomization Check, by Treatment Group												
Treatment	Control		(H) Increase		(H) Decrease		(S) Top Half		(S) Bottom Half		ANOVA	
Controls	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	F-statistic	Prob > F
Age	44.90	12.85	47.74	12.59	47.10	12.44	4429	13.09	45.64	10.83	0.83	0.51
Education	5.03	2.11	4.77	2.41	5.36	2.22	5.55	1.86	5.14	1.90	1.21	0.31
Female	0.50	0.50	0.43	0.50	0.52	0.50	0.48	0.50	0.52	0.50	0.34	0.85
Yrs for profit	17.85	15.08	14.44	12.34	16.64	12.34	12.84	12.53	12.68	10.01	1.45	0.22
Yrs public	14.94	UL22	13.47	10.55	18.22	12.81	17.27	12.80	16.46	9.83	0.96	0.43
Yrs workforce	24.45	13.72	27.69	13.02	26.57	11.86	22.28	12.36	25.15	11.07	1.65	0.16

This table provides the results of a randomization check for Experiment 1. For each treatment group, we provide the mean and standard deviation for each of six potential control variables. We ran ANOVA tests on each of the controls as a check on randomization. Of note, each of the controls appears to pass this check at the p 0.05 level.

TABLE 3
English Exam (ANOVA)

Experiment 1: English Exam				
Sample	df_between	df_within	F ratio	*p*-value
Full	4	295	10.68	0.0000
Private	4	145	4.57	0.0017
Public	4	145	7.36	0.0000

TABLE 4
English Exam: Mean Performance Assessment

Experiment I Mean Performance Assessment			
	Sample		
Vignette	*Public Sector*	*Private Sector*	*Full Sample*
Control	76.15	73.12	74.46
2% Inc (H), Upper (S)	73.21	76.6	74.4
2% Inc (H), Lower (S)	56.52*	62.97*	60.29*
2% Dec (H), Upper (S)	72.56	74.57	73.53
2% Dec (H), Lower (S)	63.86*	61.34*	62.58*

Values represent the mean of the assessed performance of "High School A" by treatment group. An * indicates those groups which are statistically distinct from the "Control" vignette at the $p < 0.05$ level.

sector and public-sector managers. We present results for the full sample with some extra discussion about the managers employed in the public sector.

For Experiment I, we conducted a one-way between-subjects ANOVA to compare the effect of historical and social comparison performance information on assessed performance of a high school in five conditions: control; increase (historical), upper half (social); increase (historical), bottom half (social); decrease (social), upper half (social); and decrease (historical), bottom half (social). Table 3 provides the one-way ANOVA results for the full sample of respondents, as well as for both the private- and public-sector respondents separately. As the table suggests, there was a statistically significant effect for the independent variable on the dependent variable at the $p < .05$ level for the five conditions in the full sample, as well as within each of the sub-samples: private- and public-sector managers.[4]

While the statistically significant findings for the treatments in the one-way ANOVA are intriguing, for the purposes of hypothesis testing we are most interested in which specific groups were statistically distinct from one another. We will present the evidence for these differences in several ways. First, Table 4 presents the means of assessed performance of "High School A" for each experimental vignette across all three samples. The different treatments (upper half (social),

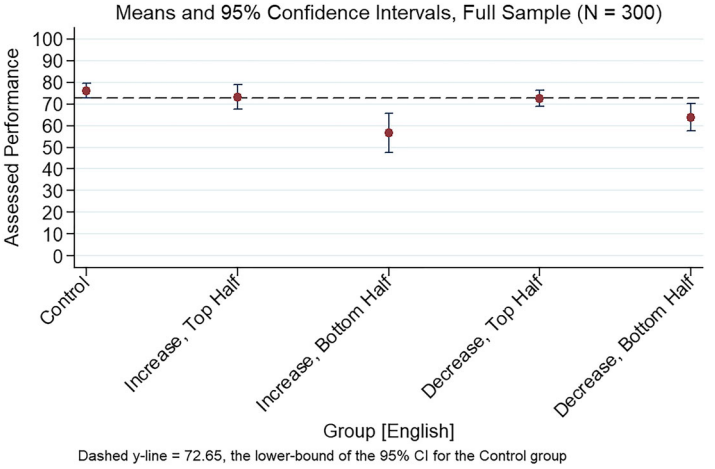

Figure 1. English exam: means and 95% CIs.

increase (historical); upper half (social), decrease (historical); etc.) are presented in the rows, and means that are statistically distinct from the control vignette are marked with an asterisk. The results suggest that respondents reacted most clearly to the negative social comparison. The assessment of performance information by that group was significantly lower than the control group, regardless of whether the historical performance was increasing or decreasing. It is important to remember here that all groups were given the same objective information about the school's performance. We do not see significant differences from the control group in the assessments of performance information in either of the groups exposed to positive social comparison, regardless of the presentation of historical information (increasing or decreasing) they received. It is also worthwhile to note that none of the differences between the two sub-sample groups (public vs. private) were statistically different from one another.

For ease of interpretation, these results are presented graphically in Figure 1. The dashed line represents the lower bound of the 95% confidence interval for the assessed performance in the control group. The 95% confidence intervals for both groups saw the "bottom half" social comparison fall below this line, suggesting that this treatment causes respondents in these groups to assess the performance of the school more negatively than the control group.

Because the independent variable was categorical, we ran three different post-hoc analyses including the Bonferroni, the Scheffe, and the Sidak. The results across all three tests are similar; in the interest of brevity, we present only the Bonferroni analysis in Table 5. The analysis suggests that the differences across groups are substantively meaningful. Across the full sample, the group which saw the historical increase and the negative social comparison had a mean response that was more than 14 points lower than the mean of the control group. This is a difference of 3.19 standard deviations. In addition, the group which saw the historical decrease and

TABLE 5
English Exam: Bonferroni Comparisons

Experiment I: English Exam: Bonferroni by Treatment Group Comparisons				
Treatment Groups	*Control*	*Increase (H), Upper Half (S)*	*Increase (H), Bottom Half (S)*	*Decrease (H), Upper Half (S)*
Increase (H), Upper Half (S)	–.054 1.000			
Increase (H), Bottom Half (S)	–14.165* 0.000	–14.112* 0.000		
Decrease (H), Upper Half (S)	–.925 1.000	–.871 1.000	13.24* 0.000	
Decrease (H), Bottom Half (S)	–11.878* 0.002	–11.825* 0.002	2.287 1.000	–10.953* 0.004

Full sample comparisons: The raw numbers in each comparison represent the difference between the column group from the row group. "H" and "S" represent historical and social treatment conditions, respectively. Numbers marked with an * represent comparisons in which the difference between the groups was statistically significant at the $p < 0.05$ level.

the negative social comparison had a mean response that was almost 12 points lower than the mean of the control group. This is a difference of 2.82 standard deviations. It is also worth noting that the results for the statistically significant differences between the respondents in the control group and respondents in the two groups which saw the negative social comparison were robust across all three sample groups: full sample, private sector, and public sector.[5]

We now move on to Experiment II. As a reminder, in this case we present respondents with either an historical or social comparison, to allow for an assessment of their independent influence on the interpretation of performance information. We again begin the analysis with a one-way between-subjects ANOVA to compare the effect of historical or social comparison performance information on the assessed performance of a high school in five conditions: control; increase (historical); decrease (historical); upper half (social); and bottom half (social). The results, presented in Table 6, suggest a statistically significant effect ($p < 0.05$) for the treatment in the full sample as well as within each of the sub-samples (private- and public-sector managers).

Again, however, our real interest is in the degree to which different reference points influence the assessment of performance information and so, we present the mean response by survey vignette in Table 7. The different groups are presented in the rows of the table and significant differences are marked by an asterisk. The findings suggest that respondents in the "bottom half" social comparison provided performance assessments that were significantly lower ($p < 0.05$) than those provided by the control group. This was true for the full sample and held within each of the two sub-samples as well. No other groups provided responses which were meaningfully different from that of the control

TABLE 6
Math Exam (Exp II): (ANOVA)

Experiment II: Math Exam				
Sample	df_between	df_within	F ratio	*p*-value
Full	4	295	10.48	0.0000
Private	4	145	6.77	0.0001
Public	4	145	4.03	0.0039

TABLE 7
Math Exam: Mean Performance Assessment

Experiment II Mean Performance Assessment			
	Sample		
Vignette	*Public Sector*	*Private Sector*	*Full Sample*
Control	76.82	75.34	76.14
2% Inc (H)	75.76	72.29	74.42
2% Decrease (H)	73.44	74.23	73.86
Upper Half (S)	75.84	80.03	78.22
Bottom Half (S)	64.5*	62.36*	63.31*

Values represent the mean of the assessed performance of "High School A" by treatment group. An * indicates those groups which are statistically distinct from the "Control" vignette at the $p < 0.05$ level.

group. For ease of interpretation, we also present the results of the mean comparison graphically in Figure 2. The dashed y line represents the lower bound of the 95% confidence interval for the assessed performance in the control group. The only group for which the 95% confidence intervals fall below this line is the "bottom half" social comparison.

This result reinforces the findings from the first experiment that negative social comparisons significantly influence how managers assess performance information. At the same time, the results of this experiment do not support our first hypothesis that managers will respond to historical reference points.

To confirm the robustness of the result, we again present a Bonferroni comparison of the control and treatment groups (Table 8). When we consider our full sample, the bottom-half social comparison is statistically distinct from the control group, as well as all the other treatment groups in this comparison. None of the other treatment groups are statistically different from the control group. In the full sample, these comparison results were robust across the Scheffe and Sidak comparisons. Additionally, across the full sample, the group which saw the "bottom half" social comparison had a mean response that was almost 13 points lower than the mean of the control group. This is a difference of 2.59 standard deviations, which again represents a substantively meaningful impact.[6]

TABLE 8
Math Exam: Bonferroni Comparisons

Experiment II: Math Exam: Bonferroni by Treatment Group Comparison				
Treatment Groups	Control	Increase (H)	Decrease (H)	Upper Half (S)
Increase (H)	–1.724 1.000			
Decrease (H)	–2.281 1.000	–.557 1.000		
Upper Half (S)	2.081 1.000	3.805 1.000	4.362 0.882	
Bottom Half (S)	–12.838* 0.000	–11.114* 0.000	–10.557* 0.000	–14.919* 0.000

Full sample comparisons: The raw numbers in each comparison represent the difference between the column group from the row group. "H" and "S" represent historical and social treatment conditions, respectively. Numbers marked with an * represent comparisons in which the difference between the groups was statistically significant at the $p < 0.05$ level.

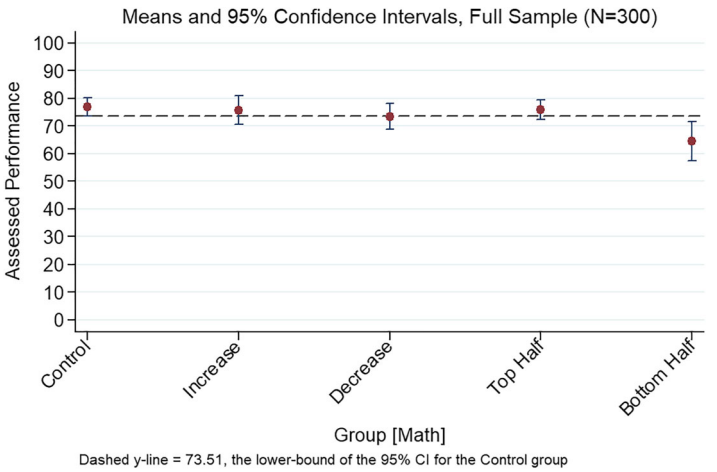

Figure 2. Math exam: means and 95% CIs.

DISCUSSION

Motivated by a growing research program in public administration on the use of performance information, we design a set of experiments to understand how reference points, or the comparison of metrics to preestablished benchmarks of acceptability, influence that process. Specifically, we explore whether comparisons of performance information against previous performance (historical comparison) or the performance of peer institutions (social comparison) influences the interpretation of that information. We build on previous work in the public administration

literature, which focuses on the use of historical and social comparisons by citizens, by testing for these reference point effects in managers.

In terms of substantive results, we find that social reference points matter. More specifically, the results suggest that comparisons with a peer significantly affect how managers interpret performance information and, consistent with our expectations, negative deviations from the social reference point matter more. Indeed, managers who were told that the performance metric they saw was in the bottom half of peer schools rated organizational performance significantly lower than the control group across both experiments. We did not find a significant effect for historical comparisons, regardless of whether performance was increasing or decreasing and regardless of whether historical shifts were paired with or administered as separate treatments from social comparisons. This result is a deviation from previous work on the interpretation of performance information by citizens and warrants further discussion, which we believe can help to illuminate the contributions of this study.

The first of these is the fact that this study investigates reference effects on the interpretation of performance information in a sample of *professional managers*. As noted earlier, this group is among the primary targets for such information and there is evidence that elites use information differently than citizens. As such, we might expect differences between the two groups, and that is exactly what we find. For example, while Olsen (2017) and Charbonneau and Van Ryzin (2015) concluded that social comparisons probably play a stronger role in citizen evaluations than do historical comparisons, both studies still found some influence for the latter. We do not replicate that result in our pool of professional managers, suggesting that this group may place less emphasis on historical comparisons than do citizens.

This result has some potentially significant implications. First, it suggests that studies of citizens may offer an incomplete picture of performance information use by professional managers and implies that the latter should be the subject of more research, despite the difficulties. Second, the results suggest that we may want to frame performance information in different ways, choosing different reference points, depending on the target group for that information.

The strong effect of social comparisons in our study draws attention to what we believe is another contribution. It is important to remember that previous comparisons of social versus historical reference points were drawn from separate experiments (Olsen 2017) or from experiments that could not accommodate all of the potential points comparisons (Charbonneau and Van Ryzin 2015). Alternatively, the design in Experiment I allows us to directly compare the influence of social and historical frames. The confirmation of the relative importance of social comparisons in that design represents another contribution to this literature.

Finally, we believe that including managers from both the public and private sectors in our subject pool represents a contribution to work on performance information use and to the longstanding debate on differences between these sectors. Interestingly, our results do not suggest consistent differences between public and private managers' uses of historical versus social reference points when interpreting performance information. In all but one case, the responses by managers from

different sectors to different types of comparisons and different directional changes in performance were statistically indistinguishable. The only significant difference we observed was in the second experiment, where private sector managers were more responsive to *positively* framed information when making social comparisons than were their public sector counterparts. The relative lack of distinction between public and private managers in processing performance information may mean that performance measurement and management systems may be more portable across sectors than previously thought.

Before concluding, we need to acknowledge some limitations of this study that suggest the need for replication and point the way forward for future research. We acknowledge that we need to be cautious regarding inferences about the power of historical comparisons because of the relatively small annual performance change (2% historical change) that we use as a treatment. As noted earlier, we chose this figure because it is a good approximation of the average performance shift that schools experience from year to year, but it may fall within the margin of indifference and, as a result, not be enough to move the manager to update any belief they may have had about the school's performance (Meier, Favero, and Zhu et al. 2015). Another potential limitation is that a 2% annual change might not adequately capture the types of change that schools may be prone to experience outside of an experimental setting (Leckie and Goldstein 2009). Future experiments will manipulate the annual change parameter to better understand the size of this margin for historical comparisons and if the margin differs by policy area.

A related concern is a potential anchoring effect from using a sliding scale to represent a performance assessment because this format for rating performance is similar to the scale in which respondents saw the performance information. We appreciate a critique from a reviewer who raised this concern. For those scholars interested in behavioral public administration, we believe it points to the value of understanding the role of information processing in the use of performance information; for example, what does a performance rating "look like" in the mind of a user? We believe that this is an area of research which could add to our understanding of performance information use.

Additionally, as noted earlier, we did not ask our respondents to assess their own performance or the performance of the organization for which they work. Future work will attempt to better tie the information provided to subjects to their own work experience and will ask for personal, as well as organizational, assessments. Finally, respondents were asked to evaluate static performance information. While this was a limitation of sample size in this study, future work will randomize the specific performance information that subjects see (see Olsen 2017).

NOTES

1. A total of 443 people started the survey, and 100 of these were removed by Qualtrics in this final screen because they did not meet our criteria.

2. https://www.doe.in.gov/assessment/istep-results
3. It is important to note that there is also a large literature on the influence of performance information on citizen satisfaction with schools (Jacobsen and Saultz 2016; Jacobsen, Saultz, and Snyder 2013; Jacobsen, Snyder, and Saultz 2014, 2015; Wilson and Piebalga 2008). While obviously important, the purpose and findings from these studies are distinct from the purpose of this study.
4. We wish to thank those who suggested this approach to present these findings during the review process.
5. A one-way ANOVA suggests that this sample suffers from unequal variance, so, as a final robustness, we also run a Kruskal-Wallis test, which is a nonparametric alternative (Hamilton 2008). The findings are essentially identical to those reported in Table 3.
6. As we did in Experiment I, we ran a Kruskal-Wallis test because of our concerns about unequal variance in the dependent variable (the assessed performance rating of High School B, given the pass rate on the Math exam). Again, these tests indicated a difference of means, with the clear distinction coming from the lower-half social comparison.

ORCID

Sean Webeck http://orcid.org/0000-0003-4050-3764
Sean Nicholson-Crotty http://orcid.org/0000-0003-2297-6939

REFERENCES

Albert, S. 1977. "Temporal Comparison Theory." *Psychological Review* 84(6):485. doi: 10.1037/0033-295X.84.6.485.

Ammons, D. N., and W. C. Rivenbark. 2008. "Factors Influencing the Use of Performance Data to Improve Municipal Services: Evidence from the North Carolina Benchmarking Project." *Public Administration Review* 68(2):304–18. doi: 10.1111/j.1540-6210.2007.00864.x.

Andersen, S. C., and M. Hjortskov. 2016. "Cognitive Biases in Performance Evaluations." *Journal of Public Administration Research and Theory* 26(4):647–62. doi: 10.1093/jopart/muv036.

Andersen, S. C., and D. P. Moynihan. 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):448–60. doi: 10.1093/jopart/muv038.

Anderson, D. M., and B. C. Edwards. 2015. "Unfulfilled Promise: Laboratory Experiments in Public Management Research." *Public Management Review* 17(10):1518–42. doi: 10.1080/14719037.2014.943272.

Baekgaard, M., and S. Serritzlew. 2016. "Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension." *Public Administration Review* 76(1):73–82. doi: 10.1111/puar.12406.

Barrows, S., M. Henderson, P. E. Peterson, and M. R. West. 2016. "Relative Performance Information and Perceptions of Public Service Quality: Evidence from American School Districts." *Journal of Public Administration Research and Theory* 26(3):571–83. doi: 10.1093/jopart/muw028.

Behn, R. D. 1995. "The Big Questions of Public Management." *Public Administration Review* 55(4):313–24. doi: 10.2307/977122.

Behn, R. D. 2003. "Why Measure Performance? Different Purposes Require Different Measures." *Public Administration Review* 63(5):586–606. doi: 10.1111/1540-6210.00322.

Berman, E., and X. Wang. 2000. "Performance Measurement in Us Counties: Capacity for Reform." *Public Administration Review* 60(5):409–20. doi: 10.1111/0033-3352.00104.

Bourdeaux, C., and G. Chikoto. 2008. "Legislative Influences on Performance Management Reform." *Public Administration Review* 68(2):253–65. doi: 10.1111/j.1540-6210.2007.00860.x.

Bouwman, R., and S. Grimmelikhuijsen. 2016. "Experimental Public Administration from 1992 to 2014: A Systematic Literature Review and Ways Forward." *International Journal of Public Sector Management* 29(2):110–31. doi: 10.1108/IJPSM-07-2015-0129.

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

Charbonneau, E., and F. Bellavance. 2012. "Blame Avoidance in Public Reporting." *Public Performance & Management Review* 35(3):399–421. doi: 10.2753/PMR1530-9576350301.

Charbonneau, É., and G. G. Van Ryzin. 2015. "Benchmarks and Citizen Judgments of Local Government Performance: Findings from a Survey Experiment." *Public Management Review* 17(2):288–304. doi: 10.1080/14719037.2013.798027.

Festinger, L. 1954. "A Theory of Social Comparison Processes." *Human Relations* 7(2): 117–40. doi: 10.1177/001872675400700202.

Freeman, R. A. 1954. "How Can Voters Understand?" *National Municipal Review* 43(3): 123–9. doi: 10.1002/ncr.4110430305.

Gaines, B. J., J. H. Kuklinski, P. J. Quirk, B. Peyton, and J. Verkuilen. 2007. "Same Facts, Different Interpretations: Partisan Motivation and Opinion on Iraq." *Journal of Politics* 69(4):957–974. doi: 10.1111/j.1468-2508.2007.00601.x.

George, B., S. Desmidt, P. Nielsen, and M. Baekgaard. 2016. "Rational Planning and Politicians' Attitudes to Spending and Reform: Replication and Extension of a Survey Experiment." Presented at the 2016 IRSPM Annual Conference

Hamilton, L. C. 2008. *Statistics with Stata: Version 10*. Belmont, CA, Brooks/Cole.

Ho, A. T.-K. 2006. "Accounting for the Value of Performance Measurement from the Perspective of Midwestern Mayors." *Journal of Public Administration Research and Theory* 16(2):217–37. doi: 10.1093/jopart/mui046.

Hvidman, U., and S. C. Andersen. 2016. "Perceptions of Public and Private Performance: Evidence from a Survey Experiment." *Public Administration Review* 76(1):111–20. doi: 10.1111/puar.12441.

Ito, T. A., J. T. Larsen, N. K. Smith, and J. T. Cacioppo. 1998. "Negative Information Weighs More Heavily on the Brain: The Negativity Bias in Evaluative Categorizations." *Journal of Personality and Social Psychology* 75(4):887. doi: 10.1037/0022-3514.75.4.887.

Jacobsen, R., and A. Saultz. 2016. "Will Performance Management Restore Citizens' Faith in Public Education?" *Public Performance & Management Review* 39(2):476–497. doi: 10.1080/15309576.2015.1108790.

Jacobsen, R., A. Saultz, and J. W. Snyder. 2013. "When Accountability Strategies Collide: Do Policy Changes That Raise Accountability Standards also Erode Public Satisfaction?" *Educational Policy* 27(2):360–389. doi: 10.1177/0895904813475712.

Jacobsen, R., J. W. Snyder, and A. Saultz. 2014. "Informing or Shaping Public Opinion? The Influence of School Accountability Data Format on Public Perceptions of School Quality." *American Journal of Education* 121(1):1–27. doi: 10.1086/678136.

Jacobsen, R., J. W. Snyder, and A. Saultz. 2015. "Understanding Satisfaction with Schools: The Role of Expectations." *Journal of Public Administration Research and Theory* 25(3): 831–848. doi: 10.1093/jopart/muu026.

James, O. 2011. "Managing Citizens' Expectations of Public Service Performance: Evidence From Observation and Experimentation in Local Government." *Public Administration* 89(4):1419–1435. doi: 10.1111/j.1467-9299.2011.01962.x.

James, O., and A. Moseley. 2014. "Does Performance Information About Public Services Affect Citizens' Perceptions, Satisfaction, and Voice Behaviour? Field Experiments With Absolute and Relative Performance Information." *Public Administration* 92(2): 493–511. doi: 10.1111/padm.12066.

James, O., and P. John. 2007. "Public Management at the Ballot Box: Performance Information and Electoral Support for Incumbent English Local Governments." *Journal of Public Administration Research and Theory* 17(4):567–580. doi: 10.1093/jopart/mul020.

James, O., S. R. Jilke and G. G. V. Ryzin, (ed). 2017. *Experiments in Public Management Research: Challenges and Contributions*. New York, Cambridge University Press.

Jerit, J., J. Barabas, and T. Bolsen. 2006. "Citizens, Knowledge, and the Information Environment." *American Journal of Political Science* 50(2):266–282. doi: 10.1111/j.1540-5907.2006.00183.x.

Jilke, S., S. Van de Walle, and S. Kim. 2016. "Generating Usable Knowledge through an Experimental Approach to Public Administration." *Public Administration Review* 76(1): 69–72. doi: 10.1111/puar.12488.

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

Kahneman, D. 2002. "Maps of Bounded Rationality: A Perspective on Intuitive Judgment and Choice." *Nobel Prize Lecture* 8:449–89.

Kroll, A. 2014. "Why Performance Information Use Varies Among Public Managers: Testing Manager-Related Explanations." *International Public Management Journal* 17(2):174–201. doi: 10.1080/10967494.2014.905409.

Kroll, A. 2015a. "Drivers of Performance Information Use: Systematic Literature Review and Directions for Future Research." *Public Performance & Management Review* 38(3): 459–86. doi: 10.1080/15309576.2015.1006469.

Kroll, A. 2015b. "Explaining the Use of Performance Information by Public Managers: A Planned-Behavior Approach." *The American Review of Public Administration* 45(2): 201–15. doi: 10.1177/0275074013486180.

Leckie, G., and H. Goldstein. 2009. "The Limitations of Using School League Tables to Inform School Choice." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172(4):835–51. doi: 10.1111/j.1467-985X.2009.00597.x.

Luskin, R. 1987. "Measuring Political Sophistication." *American Journal of Political Science* 31(4):856–899. doi: 10.2307/2111227.

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

Mintz, A., S. B. Redd, and A. Vedlitz. 2006. "Can We Generalize from Student Experiments to the Real World in Political Science, Military Affairs, and International Relations?" *Journal of Conflict Resolution* 50(5):757–776. doi: 10.1177/0022002706291052.

Moynihan, D. P. 2005. "Goal-Based Learning and the Future of Performance Management." *Public Administration Review* 65(2):203–16. doi: 10.1111/j.1540-6210.2005.00445.x.

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

Moynihan, D. P. 2010. "The Promises and Paradoxes of Performance-Based Bureaucracy." Pp. 278–302 in *The Oxford Handbook of American Bureaucracy*, edited by R. F. Durant. New York, NY, Oxford University Press.

Moynihan, D. P. 2015. "Uncovering the Circumstances of Performance Information Use Findings from an Experiment." *Public Performance & Management Review* 39(1):33–57. doi: 10.1080/15309576.2016.1071160.

Moynihan, D. P., and D. P. Hawes. 2012. "Responsiveness to Reform Values: The Influence of the Environment on Performance Information Use." *Public Administration Review* 72(s1):S95–S105. doi: 10.1111/j.1540-6210.2012.02653.x.

Moynihan, D. P., and P. W. Ingraham. 2004. "Integrative Leadership in the Public Sector a Model of Performance-Information Use." *Administration & Society* 36(4):427–53. doi: 10.1177/0095399704266748.

Moynihan, D. P., and N. Landuyt. 2009. "How Do Public Organizations Learn? Bridging Cultural and Structural Perspectives." *Public Administration Review* 69(6):1097–105. doi: 10.1111/j.1540-6210.2009.02067.x.

Moynihan, D. P., and S. Lavertu. 2012. "Does Involvement in Performance Management Routines Encourage Performance Information Use? Evaluating Gpra and Part." *Public Administration Review* 72(4):592–602. doi: 10.1111/j.1540-6210.2011.02539.x.

Moynihan, D. P., P. A. Nielsen, and A. Kroll. 2017. "Managerial Use of Performance Data by Bureaucrats and Politicians." Pp. 244–69 in *Experiments in Public Management Research: Challenges and Contributions*, edited by O. James, S. R. Jilke and G. G. V. Ryzin. New York, NY, Cambridge University Press.

Moynihan, D. P., and S. K. Pandey. 2010. "The Big Question for Performance Management: Why Do Managers Use Performance Information?" *Journal of Public Administration Research and Theory* 20(4):849–66. doi: 10.1093/jopart/muq004.

Moynihan, D. P., S. K. Pandey, and B. E. Wright. 2012a. "Prosocial Values and Performance Management Theory: Linking Perceived Social Impact and Performance Information Use." *Governance* 25(3):463–83. doi: 10.1111/j.1468-0491.2012.01583.x.

Moynihan, D. P., S. K. Pandey, and B. E. Wright. 2012b. "Setting the Table: How Transformational Leadership Fosters Performance Information Use." *Journal of Public Administration Research and Theory* 22(1):143–64. doi: 10.1093/jopart/mur024.

Mussweiler, T. 2003. "Comparison Processes in Social Judgment: Mechanisms and Consequences." *Psychological Review* 110(3):472–89. doi: 10.1037/0033-295X.110.3.472.

Mussweiler, T., and K. Epstude. 2009. "Relatively Fast! Efficiency Advantages of Comparative Thinking." *Journal of Experimental Psychology: General* 138(1):1–21. doi: 10.1037/a0014374.

Mussweiler, T., and A.-C. Posten. 2012. "Relatively Certain! Comparative Thinking Reduces Uncertainty." *Cognition* 122(2):236–40. doi: 10.1016/j.cognition.2011.10.005.

Nielsen, P. A. 2013. "Performance Information in Politics and Public Management: Impacts on Decision Making and Performance." PhD dissertation, Department of Political Science and Government. Aarhus, Denmark, Aarhus University.

Nielsen, P. A., and M. Baekgaard. 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):545–69. doi: 10.1093/jopart/mut051.

Nielsen, P. A., and D. P. Moynihan. 2017. "How Do Politicians Attribute Bureaucratic Responsibility for Performance? Negativity Bias and Interest Group Advocacy." *Journal of Public Administration Research and Theory* 27(2):269–83. doi:10.1093/jopart/muw060.

Olsen, A. L. 2013. "Leftmost-Digit-Bias in an Enumerated Public Sector? An Experiment on Citizens' Judgment of Performance Information." *Judgment and Decision Making* 8(3):365–71.

Olsen, A. L. 2014. "The Appropriate Response to Performance Information by Citizens: A Large-Scale Experiment with Local Politicians." Det Natur- og Biovidenskabelige Fakultet.

Olsen, A. L. 2015a. "Citizen (Dis) Satisfaction: An Experimental Equivalence Framing Study." *Public Administration Review* 75(3):469–78. doi: 10.1111/puar.12337.

Olsen, A. L. 2015b. "The Numerical Psychology of Performance Information: Implications for Citizens, Managers, and Policymakers." *Public Performance & Management Review* 39(1):100–15. doi: 10.1080/15309576.2016.1071167.

Olsen, A. L. 2017. "Compared to What? How Social and Historical Reference Points Affect Citizens' Performance Evaluations." *Journal of Public Administration Research and Theory* 27(4):562–80. doi: 10.1093/jopart/mux023.

Rozin, P., and E. B. Royzman. 2001. "Negativity Bias, Negativity Dominance, and Contagion." *Personality and Social Psychology Review* 5(4):296–320. doi: 10.1207/S15327957PSPR0504_2.

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

Simon, H. A. 1939. "The Administrator in Search of Statistics." *Public Management* 21: 106–9.

Simon, H. A. 1955. "A Behavioral Model of Rational Choice." *The Quarterly Journal of Economics* 69(1):99–118. doi: 10.2307/1884852.

Stritch, J. M., J. P. Mogens, and G. Taggart. 2017. "The Opportunities and Limitations of Using Mechanical Turk (MTURK) in Public Administration and Management Scholarship." *International Public Management Journal* 20(3):489–511. doi: 10.1080/10967494.2016.1276493.

Taylor, J. 2009. "Strengthening the Link between Performance Measurement and Decision Making." *Public Administration* 87(4):853–71. doi: 10.1111/j.1467-9299.2009.01788.x.

Van de Walle, S., and W. Van Dooren. 2011. "Introduction: Using Public Sector Performance Information." Pp. 1–12 in *Performance Information in the Public Sector: How It Is Used*, edited by W. Van Dooren and S. Van de Walle. New York, NY, Palgrave Macmillan.

Wilson, D. and A. Piebalga. 2008. "Performance Measures, Ranking and Parental Choice: An Analysis of the English School League Tables." *International Public Management Journal* 11(3):344–366. doi: 10.1080/10967490802301336.

Yang, K., and J. Y. Hsieh. 2007. "Managerial Effectiveness of Government Performance Measurement: Testing a Middle-Range Model." *Public Administration Review* 67(5): 861–79. doi: 10.1111/j.1540-6210.2007.00774.x.

Yockey, M. D., and S. M. Kruml. 2009. "Everything Is Relative, but Relative to What? Defining and Identifying Reference Points." *Journal of Business and Management* 15(1):95.

ABOUT THE AUTHORS

Sean Webeck (sean.webeck@nps.edu) is an assistant professor in the Defense Resources Management Institute at the Naval Postgraduate School. His research explores the role of decision making in the management of public-sector organizations. His research interests touch on behavioral public administration, human resource management, and performance management.

Sean Nicholson-Crotty (seanicho@indiana.edu) is a professor in the School of Public and Environmental Affairs and adjunct professor of political science, both at Indiana University. His research focuses on management of public organizations, diffusion of public policies among governments, and dimensions of U.S. federalism.

APPENDIX

Introduction (all respondents)	You will now be asked to respond to a question pertaining to educational performance data. For this question you are to assess the performance of an unnamed high school. For this question you will see the pass rate on a standardized **ENGLISH** exam. Administrators at the high school recently found out the results of yearly standardized testing. They want you to provide a performance assessment based upon the information you will see. Only use the information in front of you to assess each school's performance.				
Vignettes (random assignment)	*Control*	*Historical (Increase), Social (Top half)*	*Historical (Increase), Social (Bottom half)*	*Historical (Decrease), Social (Top half)*	*Historical (Decrease), Social (Bottom half)*
Vignette text	English Exam: **77**% of students in "High School A" passed their standardized English exam.	English Exam: **77**% of students in "High School A" passed their standardized English exam. This represents a 2% increase from the previous year. It also means the school was in the top half of local schools in the area (3rd out of 10).	English Exam: **77**% of students in "High School A" passed their standardized English exam. This represents a 2% increase from the previous year. It also means the school was in the bottom half of local schools in the area (7th out of 10).	English Exam: **77**% of students in "High School A" passed their standardized English exam. This represents a 2% decrease from the previous year. It also means the school was in the top half of local schools in the area (3rd out of 10).	English Exam: **77**% of students in "High School A" passed their standardized English exam. This represents a 2% decrease from the previous year. It also means the school was in the bottom half of local schools in the area (7th out of 10).
Assessment prompt (all respondents)	Assuming this is the only information available to you, use the sliding scale (0-100) to assess the overall performance of HIGH SCHOOL A over the last year:				

Figure A1. Experiment I vignette workflow.

Introduction (all respondents)	You will now be asked to respond to a question pertaining to educational performance data. For this question you are to assess the performance of an unnamed high school. For this question you will see the pass rate on a standardized **MATH** exam. Administrators at the high school recently found out the results of yearly standardized testing. They want you to provide a performance assessment based upon the information you will see. Only use the information in front of you to assess each school's performance.				
Vignettes (random assignment)	*Control*	*Historical (Increase)*	*Historical (Decrease)*	*Social (Top half)*	*Social (Bottom half)*
Vignette text	Math Exam: **79**% of students in "High School B" passed their standardized Math exam.	Math Exam: **79**% of students in "High School B" passed their standardized Math exam. This represents a 2% increase from the previous year.	Math Exam: **79**% of students in "High School B" passed their standardized Math exam. This represents a 2% decrease from the previous year.	Math Exam: **79**% of students in "High School B" passed their standardized Math exam. This means the school was in the top half of local schools in the area.	Math Exam: **79**% of students in "High School B" passed their standardized Math exam. This means the school was in the bottom half of local schools in the area.
Assessment prompt (all respondents)	Assuming this is the only information available to you, use the sliding scale (0-100) to assess the overall performance of HIGH SCHOOL B over the last year:				

Figure A2. Experiment II vignette workflow.

TABLE A1
Impact of Comparison Frames on Performance Assessment with Pre-Treatment Covariate

Regressions with Control Variables

	English Exam (Experiment I)	Math Exam (Experiment II)
Treatment 1	0.23 (3.08)	–1.34 (2.45)
Treatment 2	–12.89*** (2.99)	–2.42 (2.48)
Treatment 3	–0.39 (3.00)	1.62 (2.50)
Treatment 4	–11.61*** (3.05)	–12.08*** (2.48)
Age	0.09 (.18)	–0.01 (0.15)
Education	–0.09 (.18)	0.68 (.40)
Female	0.52 (2.01)	0.60 (1.66)
Public Sector	–1.10 (1.97)	–0.36 (1.62)
Years in Workforce	–0.24 (.18)	–0.04 (.15)
Constant	77.36*** (5.37)	74.04*** (443)
N	297	297
R-squared	0.1346	0.1255

The outputs of this table provide the coefficients and standard errors from a regression of control variables in addition to the IV of interest (treatment group). The DV is the performance assessment individuals responded for each experiment.

***Significant at the 0.01 level.

Exp I. Treatment 1: H (increase), S (upper half); Treatment 2: H (increase), S (lower half); Treatment 3: H (decrease), S (upper half); Treatment 4: H (decrease), S (lower half).

Exp II. Treatment 1: H (increase); Treatment 2: H (decrease): Treatment 3: S (upper half); Treatment 4: S (lower half).