Bureaucratic Decision Making:

The Role of Historical and Social Comparisons in the Use of Performance Information

Sean Webeck

Sean Nicholson-Crotty

Indiana University

School of Public and Environmental Affairs

Department of Political Science

American Political Science Association

Annual Meeting

San Francisco, CA

August 31, 2017

Abstract: In this study, we engage the literature concerned with how people process performance information. We adopt a psychological approach which suggests that historical and social comparisons likely influence the way they interpret organizational performance, but assess those interpretations in a subject pool of professional managers, rather than citizens, which has been the norm in this literature. Evidence from two experiments suggests that social comparisons, especially when framed as poor performance relative to peer or competitor organizations, have the largest influence on the way managers process performance information. This supports results from previous work in this area. However, historical comparisons do not have a significant effect on the interpretation of performance information for this group, which represents a meaningful departure from studies using citizens. These results contribute to a growing literature in public management on the use of performance information and the role of psychological factors in helping us to understand how individuals use this information.

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, p. 33). Kroll (2015a) 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 bureaucrats *process* performance information (note, as exceptions, Andersen & Moynihan, 2016; Kroll, 2015b).

This paper contributes to this understanding by engaging the literature on relative performance evaluation. Research suggests that individuals do not judge information about performance in absolute terms, but in reference to how they performed previously or how social peers are performing currently. To date, however, scholars have focused primarily on the e historical and social comparisons by ordinary citizens. For a variety of reasons, these results may not offer an accurate picture of the ways in which professional bureaucrats use these heuristics when judging performance information.

We offer results from two experiments designed to distinguish the relative importance of historical and social comparisons of performance survey of 300 public and private sector managers. In line with previous research on citizens (Charbonneau & 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 do not find evidence that historical comparisons have a significant

influence on the way in which professional managers interpret performance information. This provides empirical evidence in support of a behavioral—i.e., psychological—approach to understanding the use performance information, but also suggests that findings from studies of citizens may not generalize perfectly to studies of managers. Finally, the results provides potentially important insights regarding the design of systems designed to provide performance information to this group and about the ways in which performance information should be framed for different audiences.

In what follows we present an argument that addresses our research question: what role do historical and social comparisons play in how bureaucrats use and interpret performance information? First, we review scholarly literature pertinent to the question. Then, we discuss the advantages of an experimental approach and explain our experimental research design. We then describe two experiments and present their results. We conclude with a discussion and some practical and scholarly implications of this research.

PERFORMANCE INFORMATION

Bureaucrats and the "Use" of Performance Information

From a both academic and practical perspective, one of the compelling aspects of performance managements systems is that they generate, at least ostensibly, an objective assessment of how an organization is doing because. This is because the information these systems generate—performance information—is intended to be "systematic" information (Poul Aaes Nielsen, 2013). And yet, *who* or *what organization* assesses performance metrics potentially influences the how the information is interpreted (Moynihan, 2008). While scholars study the use of performance information across three groups of "end users"—citizens, managers, and

politicians (Van de Walle & Van Dooren, 2011), bureaucrats represent the primary end users of performance information. As a result, we suggest their decisions about how to use performance information should garner the most attention from scholars working in this area.

In one of the first enumerated lists on the subject of how bureaucrats use performance information, 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 (ibid, p. 588).

Taking a slightly different approach, rather than looking at specific actions, Moynihan (2010) suggests there are four strategies public managers can employ when using performance information: passive, political, perverse, and purposeful. Yet, for Moynihan, the existence of other interested actors (e.g. political principals and citizens) makes performance data more subjective than they are often argued to be. He argues this plays an important role in how we should understand performance information use.

A recent systematic literature review offers another set of factors which help us understand why bureaucrats use performance information (Kroll, 2015a). Over the last 15 years in "a highly relevant and fast-growing research area" (ibid., 460), research consistently shows six factors commonly drive the use of performance information among bureaucrats: measurement system maturity (for examples see Berman & Wang, 2000; Ho, 2006; Taylor, 2009), stakeholder involvement (for examples see Bourdeaux & Chikoto, 2008; Ho, 2006; Moynihan & Pandey, 2010), leadership support (for examples see Moynihan & Ingraham, 2004; Moynihan & Lavertu, 2012; Yang & Hsieh, 2007), support capacity (for examples see Berman & Wang, 2000; Julnes & Holzer, 2001; Moynihan & Hawes, 2012), innovative culture (for examples see Moynihan, 2005;

Moynihan & Pandey, 2010; Moynihan, Pandey, & Wright, 2012b), and goal clarity (for examples see Moynihan & Landuyt, 2009; Moynihan, Pandey, & Wright, 2012a; Moynihan et al., 2012b). For our purposes, the most important takeaway from this review of previous work on performance information use is that the vast majority of studies use organizational behavior, organizational theory, or a combination of these approaches to explain the phenomenon. In other words, they are focused on the ways in which the organizational context in which performance measurement systems are embedded have the greatest influence on how the information from those systems gets used.

A more recent trend seems to push on these previous boundaries. Recent scholarship has warned that previous work has ignored the ways in which *individual* level behavior might produce systematic variation in the use of performance information among bureaucrats (Kroll, 2015a; Moynihan, Nielsen, & Kroll, 2017). For example, several studies over the past decade suggest psychology may play a role in helping us understand how bureaucrats use performance information (Andersen & Moynihan, 2016; Kroll, 2015b; Moynihan, 2008, 2015; Poul A Nielsen, 2013; Salge, 2011). 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 (Andersen & Hjortskov, 2015; Baekgaard & Serritzlew, 2016; Barrows, Henderson, Peterson, & West, 2016; Hvidman & Andersen, 2016; Olsen, 2013, 2015a, 2017), but other work has looked at how politicians (George, Desmidt, Nielsen, & Baekgaard, 2016; Poul A. Nielsen & Baekgaard, 2015; Poul A Nielsen & Moynihan, 2016; Olsen, 2014) use performance information as well. There are two important takeaways from these studies. First, experimental methods are a useful method to develop our understanding of performance information use across a variety of political actors

(Anderson & Edwards, 2015; Bouwman & Grimmelikhuijsen, 2016; James, Jilke, & Ryzin, 2017; Jilke, Van de Walle, & Kim, 2016). Second, when it comes to performance metrics, these studies suggest that depending upon the circumstances, individuals suffer from a collection of cognitive biases and utilize a number of heuristics when responding to (ostensibly objective) performance information.

Reference Points

This paper joins this growing discussion by focusing on the ways in which individual managers *interpret* performance data. Specifically, it focuses on the ways in which they respond to performance information in the presence of important reference points. 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. This idea of comparisons raises yet again the question of the objectivity of numerical performance information; or, as Simon once said, "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, as cited in Olsen [2015b]).

In the words of another Nobel Laureate, perception is "*reference-dependent*" (Kahneman, 2002, 459, emphasis in original). "Reference points are 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 & Kruml, 2009, 97). Previous scholarship points to the idea that limitations in human processing—as serial processors we are boundedly rational—constrain our ability to accept, hold, and process information (Freeman, 1954; Simon, 1955). Thus, in line with the research program on behavioral public administration, reference points potentially

represent cognitive heuristics in making evaluative judgments about performance information (Mussweiler & Epstude, 2009; Mussweiler & Posten, 2012).

Recent work on citizen evaluations of public organization performance provides some initial evidence into the value of looking at bureaucratic responses to historical and social performance reference points (Olsen, 2015b). Following previous research on reference points, we argue that reference points can play a significant role in how bureaucrats use performance information. While the logic of performance management systems suggests they generate objective performance metrics, we believe the use of reference points means that performance information is subjectively interpreted in a highly contextualized manner.

There are a variety of reference points that individuals employ when making decisions. For example, when judging gains or losses, people regularly use the status quo or some aspirational level in order to judge what is "acceptable." When judging performance, the literature suggests that individuals use both historical and social comparisons to determine whether a given level is good or bad. Historical reference points allow compare the performance of an organization to the previous performance of the same organization. For example, we might compare a school's performance metrics to those same metrics from one or more years ago. Past performance provides a nice status-quo against which individuals can easily assess change in. Since we reference the same organization we might assume that many of the factors influencing the organization and the individuals working inside it might remain stable over time. Following Olsen (2017), we hypothesize:

> *H_1: (Historical) – 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.*

In addition to historical reference points, social reference points provide another frame with which individuals compare performance. Under this approach, individuals compare the performance of their organization against the performance of other, comparable organizations. 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 longstanding 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 rating an individual gives to a public organization. Interestingly, they also found that individuals seem to give more weight to social rather than historical comparisons when using performance information. This leads to the following expectations:

> *H_2: (Social) – Providing information about better (worse) performance of other, comparable, organizations will lower (raise) a manager's assessment of an organization's current performance.*

> *H_3: (Comparison) – Managers will react more strongly to social rather than historical comparisons.*

Negativity bias and the use of performance information.

In order to accurately assess the use of historical and reference points in the interpretation of performance information, it is also important to acknowledge that people respond differently to information depending on how it is framed. The 'negativity bias,'—the idea that people

respond more strongly to negative information than to comparable information when it is framed in a positive way—is a strongly supported empirical finding (T. A. Ito, Larsen, Smith, & Cacioppo, 1998). When it comes to performance information, some recent studies from Denmark suggest evidence that citizens exhibit a negativity bias in the use of performance information. In one study, Olsen (2015a) shows that citizens evaluate public organizations differently if they believe the organization has a 90% satisfaction rate or a 10% dissatisfaction rate. In another he demonstrates that both historical and social comparisons seem to draw out the negativity bias (2017). Olsen's work looks at citizen evaluations, but the 'universality' of the negativity bias suggests we might see this bias in bureaucrats as well (T. Ito & Cacioppo, 2005; T. A. Ito et al., 1998). For this reason, we expect bureaucrats will provide lower interpretation scores for performance information that is framed negatively compared to when it is framed positively, even when the metrics being compared are qualitatively similar and regardless of whether they are making social or historical comparisons. Stated formally,

> H_4: *Managers will indicate lower levels of performance when presented with negatively (rather than positively) framed performance metrics in both social and historical comparisons.*

EXPERIMENTAL DESIGN

In this section we will discuss the survey instrument used to collect the experimental, the two experiments respondents saw, and the panel of subjects.

Background on Survey Design and Data Collection

Data for the two experiments come from a Qualtrics panel collected during May of 2017. Qualtrics screened and provided the respondents for the survey. We provided a stipulation that

respondents were managers in their organization. Unfortunately, we don't have a breakdown by job. The total sample size is 300, with 150 coming from the private-sector and 150 from the public-sector. Our sample includes managers from both sectors because another experiment in the survey required this demographic breakdown. But, in this paper 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.

It is worth noting the trade-offs we contemplated before proceeding with our survey. We wanted to run our experiments on managers. This decision played a large role into the trade-offs we made. Olsen (2017) ran a similar set of experiments in Danish citizens in which he had 3,443 respondents for both of his surveys. But, since we wanted managers instead of citizens we needed to pay more (Qualtrics quoted us $25 per manager response). This was also more expensive than if we had used a self-managed online survey platform (e.g., MTurk, TurkPrime, or Prolific) to run our survey. We ended up choosing Qualtrics to host the survey because we felt this would give us a better-quality respondent, on average. If we had gone with one of the other providers there may have been difficulty in determining the extent to which our respondents were in fact managers had we offered the study through one of the self-managed platforms. Because of our budget, the decision to run the survey through Qualtrics constrained our sample size. Results from a pilot of our survey also played into our decision to go with 300 respondents. Also, since some of our other experiments required us to look at both private- and public-sector managers we needed to strike a balance between finding significant findings and who we wanted to get in our sample. This is how we ended up with 150 managers from each sector.

Below, in Table 1, we provide the descriptive statistics for the entire sample (i.e., both samples). With an average age of nearly 46 and an average of more than 25 years in the workforce,

this group of respondents clearly has a significant amount of life and work experience. We can also see that just under half are female. There were 152 males and 145 females who responded to the survey. Three indicated "prefer not to say" on the question of gender. The mean of education for the entire sample is 5.16, which puts the average education at just more than a Bachelor's degree. The average level of education was 4.76 for the private-sector managers and 5.55 for the public-sector managers. In the sample, there were 54 Master's degrees (18 private and 36 public), 41 professional degrees (19 private and 22 public), and 8 PhD's (1 private and 7 public).

Table 1 - Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
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	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

EXPERIMENTS

One of the challenges with experiments is our ability to generalize the findings outside of the experimental setting. Since the substantive policy area is education—specifically, passing rates for standardized test scores—we decided to model our performance information off real test score data to make the experimental vignettes seem plausible. To do this we utilized publicly available data for the pass rates for standardized tests (English and Math) in a Midwestern state in the U.S. We did this by averaging the pass rates for schools in the state for both English and Math standardized exams. Since our experiments also 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).

Since we delivered both experiments in the same survey and they were very similar we took steps to reduce the possibility of respondent bias. The first step we took was to separate the two experiments in the order in which respondents saw them in our survey. 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 so we believe this was an adequate way to address this concern. Second, one 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., 3rd out of 10). This omission was deliberate and an attempt to help respondents not conflate the two experiments despite their similarities.

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 effects when both historical and social comparison information were presented together. We felt 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 could compare the primacy of each comparison as well as the strength of the positive and negative versions of each comparison as well. Of course, we were also able to compare performance assessments against the control group too.

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 that the school ranked third or seventh out of ten 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 101-point sliding scale (0-100).

As an example, someone in the group which 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 then have the opportunity to rate the performance of the school with the sliding scale.

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 *Experiment II* 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 which 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

In this section, we discuss the results of the two experiments. Since the Qualtrics sample had both private-sector and public-sector managers we present the information for the public-sector managers we present both but will focus our analysis on the public-sector managers.

Table 2 shows the randomization outcomes from Experiment I. The reason there is some variation in the number of respondents assigned to each vignette is that we randomized our assignments at the level of the survey and not by sector of employment. Thus, we see there is less variation in the number of respondents in each of the five groups across the whole experiment than

by sector. But, since we're looking at mean responses in the performance assessment respondents gave, we are willing to work with the randomization process we employed in this survey.

Table 2 - Experiment I Respondents by Sector

	Sector		
Vignette	Public	Private	Total
Control	26	33	59
2% Inc (H), Upper (S)	37	20	57
2% Inc (H), Lower (S)	27	38	65
2% Dec (H), Upper (S)	32	30	62
2% Dec (H), Lower (S)	28	29	57

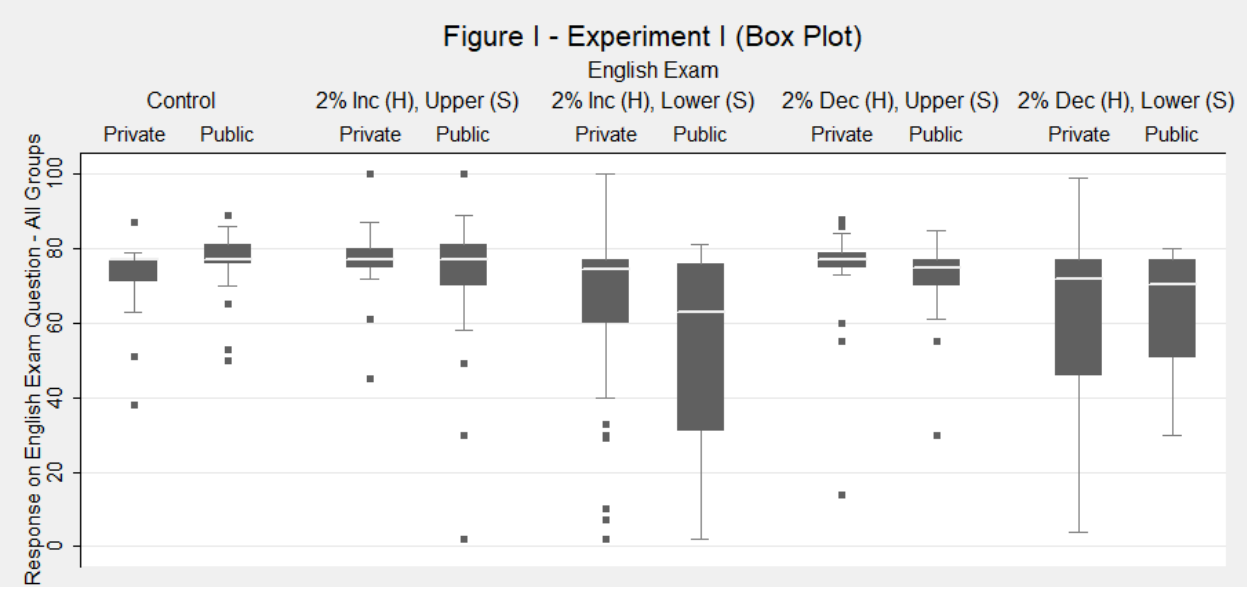
Table 3 shows the experimental results for Experiment I. Even on just a visual assessment, a few things stand out. First, on average, those who saw the lower *social* comparison information rated the performance of the school lower than the respective control groups as well as groups which saw the upper (top-half) social comparison. Empirical analysis supports this finding. A one-way ANOVA suggests the means are not equal but that this sample also suffers from unequal variance. For this reason, we run a Kruskal-Wallis test, which is a nonparametric alternative to our one-way ANOVA (Hamilton, 2008, p. 165). With the Kruskal-Wallis test we still find evidence that the means of the groups are distinct. This supports our claim of the strength of the social comparison in how bureaucrats process performance information.

It does not appear that the historical comparisons mattered for respondents in this experiment. Indeed, all the significant movement in response to performance information comes from the social comparisons. The only place where we see a difference in the historical comparisons is among public-sector managers who saw the lower frame of the social comparison. But, the difference for the historical comparison is in the opposite direction from what we would expect. We also see there is really no difference between the historical comparison and the control

group. These findings suggest the interpretations of managers are less susceptible to historical performance information but strongly responsive to social comparison data.

Table 3 - Experiment I Results			
Sector (Control)	Historical	Social	
		Upper	Lower
Public	Increase	73.21	56.52
76.15	Decrease	72.56	63.86
Private	Increase	76.6	62.97
73.12	Decrease	74.57	61.34

Further evidence of the social comparison differences comes in Figure I, where we use a box plot to graphically display the results. The plots show a clear drop in the performance appraisals of those respondents who saw the lower-performing social comparison information (i.e., that "High School A" performed worse than peer schools on the standardized English exam), but not for those that saw lower historical comparisons. This provides clear support for Hypothesis 2, but fails to provide support for our first hypothesis. Additionally, the results do not provide evidence of generalized negativity bias or support for our fourth assertion that negatively framed data would produce a larger change in assessments than positively framed information, regardless of the type of comparison.



We now move on to Experiment II. Table 4 shows the breakdown of respondents by group and by sector. Again, since we did not randomize by sector we are left with some differences across the groups in terms of the number of respondents. But, the findings seem clear that this did not significantly influence the outcome.

Table 4 - Experiment II Respondents by Sector

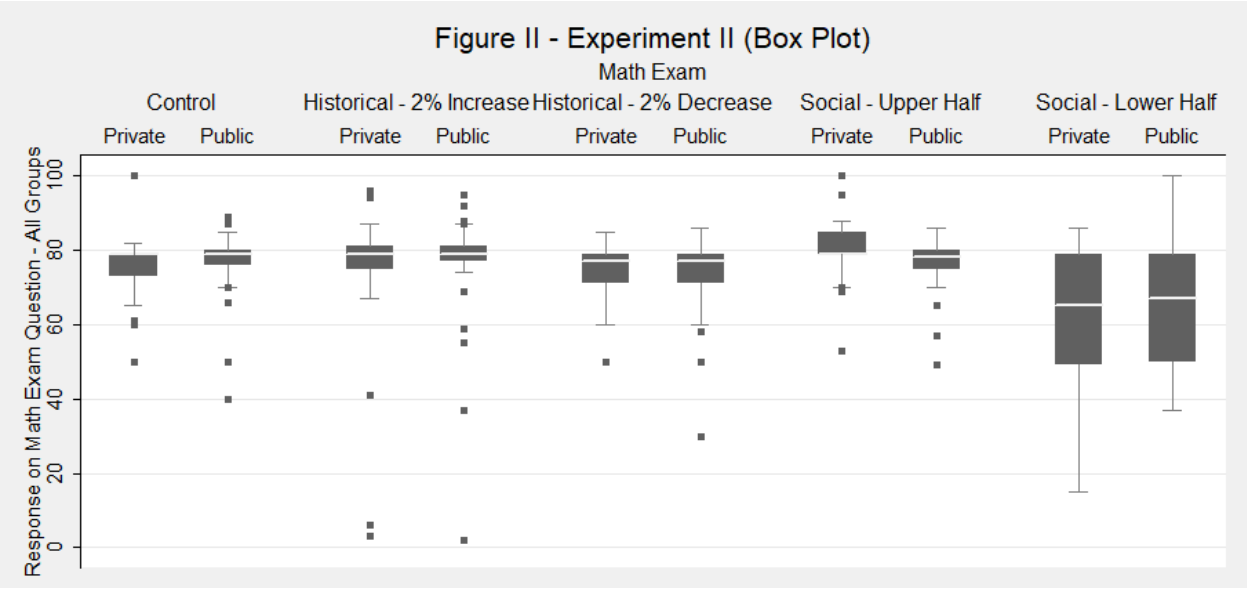
	Public	Private
Control	34	29
2% Increase	38	24
2% Decrease	27	31
Social - Upper half	25	33
Social - Lower half	26	33

Just as in Experiment I, our clear finding is that lower social comparisons significantly influence the way managers process performance information. Table 5 shows this, numerically, while Figure 2 shows this graphically. As we did in Experiment I, we ran a one-way ANOVA and then a Kruskal-Wallis test because of our concerns about unequal variance in the dependent variable (the assessed performance rating of High School B give 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. As in Experiment I, there is limited evidence of generalized negativity bias with historical responses show a negligible difference between the increase and decrease groups. Though when considering the two experiments together, it is important to note that the movement in subject responses for social comparisons were almost exclusively restricted to those instances when performance was decreasing. This does suggest that negatively framed information has a larger impact than positively framed information, and thus a negativity bias, at least in those cases where subjects are making social comparisons.

Table 5 - Experiment II Results

	Public	Private
Control	76.82	75.34
2% Increase	75.76	72.29
2% Decrease	73.44	74.23
Social - Upper half	75.84	80.03
Social - Lower half	64.5	62.36



DISCUSSION

We now turn to a discussion of our results and the implications for practice and future scholarship. Again, we built our paper around a growing research program in public administration on the use of performance information. Specifically, we were interested in understanding how managers, from both the public and private sectors, interpret performance information in the context of other signals, or reference points. We build off previous work in the public administration literature on the role of historical and social comparisons in the interpretation of performance information among citizens by testing these questions among managers. We feel that the focus on professional managers offers a significant contribution to this literature, as does the examination of both public and private sector managers, and the design of two novel experiments that allow direct comparison of the relative power of social versus historical frames.

We find that social reference points, especially data that signal an organization is underperforming vis-à-vis peer or competitor organizations, significantly affect how managers interpret performance information. We did not, however, find a significant effect for historical comparisons, regardless of whether performance was increasing or decreasing. Comparing these results with recent work on citizens' evaluations of performance highlight some interesting discrepancies. 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 potentially significant implications for the ways in which performance information should be framed, depending on the target group for

that information. Understanding that managers may process performance information differently than citizens is also important one of the primary goals if performance management systems is to get better data and information in the hands of the people *managing* public organizations.

It is also 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 the potential points comparisons (Charbonneau and Van Ryzin 2017). We believe, therefore, that the confirmation of the relative importance of social comparisons in a simple and controlled experiment which allows direct evaluation of that frame against an historical comparison 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 both 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 and point the way forward for future research. First, we did not ask our respondents to assess "real" performance

information and we did not ask them 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. Second, respondents were asked to evaluate static performance information. A better approach would probably be more like what Olsen (2017) did in which he randomized the performance information citizens saw. Of course, this speaks to the tension of sample size and cost when studying bureaucrats. Nonetheless, future scholarship in this area should focus on the magnitude of difference between a performance metric and a comparative piece of information. Finally, our experiments came out of one survey, raising the possibility of intra-survey bias between the two experiments.

References

Albert, S. (1977). Temporal comparison theory. *Psychological Review, 84*(6), 485.

Andersen, S. C., & Hjortskov, M. (2015). Cognitive Biases in Performance Evaluations. *Journal of Public Administration Research and Theory, 26*(4), 647-662. doi:doi: 10.1093/jopart/muv036

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*, 1-13. doi:doi:10.1093/jopart/muv038

Anderson, D. M., & Edwards, B. C. (2015). Unfulfilled Promise: Laboratory experiments in public management research. *Public Management Review, 17*(10), 1518-1542.

Baekgaard, M., & Serritzlew, S. (2016). Interpreting performance information: Motivated reasoning or unbiased comprehension. *Public Administration Review, 76*(1), 73-82.

Barrows, S., Henderson, M., Peterson, P. E., & West, M. R. (2016). Relative Performance Information and Perceptions of Public Service Quality: Evidence From American School Districts. *Journal of Public Administration Research and Theory*, muw028.

Behn, R. D. (1995). The big questions of public management. *Public Administration Review*, 313-324.

Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review, 63*(5), 586-606.

Berman, E., & Wang, X. (2000). Performance measurement in US counties: Capacity for reform. *Public Administration Review, 60*(5), 409-420.

Bourdeaux, C., & Chikoto, G. (2008). Legislative influences on performance management reform. *Public Administration Review, 68*(2), 253-265.

Bouwman, R., & Grimmelikhuijsen, S. (2016). Experimental public administration from 1992 to 2014: A systematic literature review and ways forward. *International Journal of Public Sector Management, 29*(2), 110-131.

Charbonneau, É., & Van Ryzin, G. G. (2015). Benchmarks and citizen judgments of local government performance: Findings from a survey experiment. *Public Management Review, 17*(2), 288-304.

Festinger, L. (1954). A theory of social comparison processes. *Human relations, 7*(2), 117-140.

Freeman, R. A. (1954). How can voters understand? *National Civic Review, 43*(3), 123-129.

George, B., Desmidt, S., Nielsen, P., & Baekgaard, M. (2016). *Rational planning and politicians' attitudes to spending and reform: replication and extension of a survey experiment.* Paper 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), 217-237.

Hvidman, U., & Andersen, S. C. (2016). Perceptions of Public and Private Performance: Evidence from a Survey Experiment. *Public Administration Review, 76*(1), 111-120.

Ito, T., & Cacioppo, J. (2005). Variations on a human universal: Individual differences in positivity offset and negativity bias. *Cognition & Emotion, 19*(1), 1-26.

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. *Journal of personality and social psychology, 75*(4), 887.

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

Jilke, S., Van de Walle, S., & Kim, S. (2016). Generating usable knowledge through an experimental approach to public administration. *Public Administration Review, 76*(1), 69-72.

Julnes, P. d. L., & Holzer, M. (2001). Promoting the utilization of performance measures in public organizations: An empirical study of factors affecting adoption and implementation. *Public Administration Review, 61*(6), 693-708.
Kahneman, D. (2002). Maps of bounded rationality: A perspective on intuitive judgment and choice *Nobel prize lecture* (Vol. 8, pp. 449-489).
Kroll, A. (2015a). Drivers of performance information use: systematic literature review and directions for future research. *Public Performance & Management Review, 38*(3), 459-486.
Kroll, A. (2015b). Explaining the Use of Performance Information by Public Managers: A planned-behavior approach. *The American Review of Public Administration*, 201-215.
Moynihan, D. P. (2005). Goal-based learning and the future of performance management. *Public Administration Review, 65*(2), 203-216.
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. In R. F. Durant (Ed.), *The Oxford Handbook of American Bureaucracy*. New York: 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.
Moynihan, D. P., & Hawes, D. P. (2012). Responsiveness to reform values: The influence of the environment on performance information use. *Public Administration Review, 72*(s1), S95-S105.
Moynihan, D. P., & Ingraham, P. W. (2004). Integrative Leadership in the Public Sector A Model of Performance-Information Use. *Administration & society, 36*(4), 427-453.
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., & Lavertu, S. (2012). Does involvement in performance management routines encourage performance information use? Evaluating GPRA and PART. *Public Administration Review, 72*(4), 592-602.
Moynihan, D. P., Nielsen, P. A., & Kroll, A. (2017). Managerial Use of Performance Data by Bureaucrats and Politicians. In O. James, S. R. Jilke, & G. G. V. Ryzin (Eds.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 244-269). New York: Cambridge University Press.
Moynihan, D. P., & Pandey, S. K. (2010). The big question for performance management: Why do managers use performance information? *Journal of Public Administration Research and Theory*, muq004.
Moynihan, D. P., Pandey, S. K., & Wright, B. E. (2012a). Prosocial values and performance management theory: Linking perceived social impact and performance information use. *Governance, 25*(3), 463-483.
Moynihan, D. P., Pandey, S. K., & Wright, B. E. (2012b). Setting the table: How transformational leadership fosters performance information use. *Journal of Public Administration Research and Theory, 22*(1), 143-164.
Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110*(3), 472-489.
Mussweiler, T., & Epstude, K. (2009). Relatively fast! Efficiency advantages of comparative thinking. *Journal of Experimental Psychology: General, 138*(1), 1-21.
Mussweiler, T., & Posten, A.-C. (2012). Relatively certain! Comparative thinking reduces uncertainty. *Cognition, 122*(2), 236-240.
Nielsen, P. A. (2013). *Performance Information in Politics and Public Management: Impacts on Decision Making and Performance.* (PhD), Aarhus University, Aarhus, Denmark.

Nielsen, P. A. (2013). Performance Management, Managerial Authority, and Public Service Performance. *Journal of Public Administration Research and Theory, 24*(2), 431-458. doi:doi: 10.1093/jopart/mut025

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. doi:10.1093/jopart/mut051

Nielsen, P. A., & Moynihan, D. P. (2016). How Do Politicians Attribute Bureaucratic Responsibility for Performance? Negativity Bias and Interest Group Advocacy. *Journal of Public Administration Research and Theory*.

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.

Olsen, A. L. (2014). *The Appropriate Response to Performance Information by Citizens: A Large-scale Experiment with Local Politicians*.

Olsen, A. L. (2015a). Citizen (dis) satisfaction: An experimental equivalence framing study. *Public Administration Review, 75*(3), 469-478.

Olsen, A. L. (2015b). The Numerical Psychology of Performance Information: Implications for Citizens, Managers, and Policymakers. *Public Performance & Management Review, 39*(1), 100-115.

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

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.

Simon, H. A. (1939). The administrator in search of statistics. *Public Management, 21*, 106-109.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics, 69*(1), 99-118.

Taylor, J. (2009). Strengthening the link between performance measurement and decision making. *Public Administration, 87*(4), 853-871.

Van de Walle, S., & Van Dooren, W. (2011). Introduction: Using Public Sector Performance Information. In W. Van Dooren & S. Van de Walle (Eds.), *Performance Information in the Public Sector: How it is Used* (pp. 1-12). New York: Palgrave Macmillan.

Yang, K., & Hsieh, J. Y. (2007). Managerial effectiveness of government performance measurement: testing a middle-range model. *Public Administration Review, 67*(5), 861-879.

Yockey, M. D., & Kruml, S. M. (2009). Everything is relative, but relative to what? Defining and identifying reference points. *Journal of Business and Management, 15*(1), 95.