International Public Management Journal

ISSN: 1096-7494 (Print) 1559-3169 (Online) Journal homepage: https://www.tandfonline.com/loi/upmj20

Are managers susceptible to framing effects? An experimental study of professional judgment of performance metrics

Javier Fuenzalida, Gregg G. Van Ryzin & Asmus Leth Olsen

To cite this article: Javier Fuenzalida, Gregg G. Van Ryzin & Asmus Leth Olsen (2020): Are managers susceptible to framing effects? An experimental study of professional judgment of performance metrics, International Public Management Journal, DOI: 10.1080/10967494.2020.1752338

To link to this article: https://doi.org/10.1080/10967494.2020.1752338

INTERNATIONAL PUBLIC MANAGEMENT JOURNAL
https://doi.org/10.1080/10967494.2020.1752338

Are managers susceptible to framing effects? An experimental study of professional judgment of performance metrics

Javier Fuenzalida[a], Gregg G. Van Ryzin[b], and Asmus Leth Olsen[c]

[a]University of Chile; [b]Rutgers University; [c]University of Copenhagen

ABSTRACT

Evidence suggests that citizens evaluate government performance differently when equivalent performance information is presented with either a positive or negative framing—but do experienced public managers also suffer from this framing effect? To address this question, we conducted an experiment with 191 public service professionals in the U.S. in which we experimentally varied the framing of performance information about customer satisfaction, job satisfaction and goal achievement for various federal government agencies. Our findings show that public service professionals—just like ordinary citizens—are susceptible to framing effects. Specifically, they tend to evaluate federal agency performance more negatively when percentages of "job dissatisfaction" and "targets not met" were presented, as opposed to logically equivalent percentages of "job satisfaction" and "targets met." The pattern is the same for "customer dissatisfaction" versus "customer satisfaction" rates, although the results are not statistically significant. These findings provide a deeper understanding about the use of government performance information, as well as how such information is comprehended and perhaps misunderstood by decision makers.

ARTICLE HISTORY
Received 21 July 2018
Accepted 2 April 2020

Introduction

A growing emphasis on performance management has emerged around the globe during the last two decades. Various public management reforms have fostered the measurement and reporting of government performance, and a large body of research and expert advice has advocated this agenda (Holmes and Shand 1995; Van Dooren, Bouckaert, and Halligan 2015). A key assumption of this performance movement holds that the tracking of metrics will help public managers make better informed decisions and can help them improve organizational performance (Poister, Aristigueta, and Hall 2014). But the performance movement entails several critical yet largely untested assumptions. One is the belief that managers indeed use such metrics, which some studies suggest happens only occasionally and is contingent on factors like their ownership and perceived social impact by officials, and their link to strategic goals, budget allocations and learning styles of agencies and individuals (Ammons and Rivenbark 2008; Kroll 2014; Kroll and Vogel 2014; Moynihan, Pandey, and Wright 2012; Taylor 2011). The demands for external accountability faced by public organizations are also a factor, as well as the level of available resources and technical expertise (de Lancer Julnes and Holzer 2001; Taylor 2009). A second key assumption is that performance information is objectively and appropriately interpreted by decision makers, an idea also questioned by several authors (Moynihan 2008; Van Dooren and Van de Walle 2008).

CONTACT Javier Fuenzalida jafuenza@dii.uchile.cl Centre for Public Systems and Institute of Public Affairs, University of Chile, Santiago, Chile.

Although some attention has been paid to how performance information should be reported (Hatry 2006), relatively little is known about how the presentation of performance information influences public managers' decision making. This gap in knowledge is important because psychology and decision science have shown that decisions can be influenced by how information is presented or framed (Tversky and Kahneman 1981). Understanding these framing effects is thus an important consideration in the design and implementation of performance measurement and reporting systems. Without an understanding of framing effects, the design and use of performance measurement and reporting systems might result in misguided policy and management decisions, detracting from the promise of the performance movement in public administration (Behn 2002).

Moreover, there is recent evidence to suggest that framing effects matter in the interpretation of government performance information (Boyne et al. 2009; James 2011; James and Moseley 2014). In particular, previous work by Olsen (2015a) demonstrated that presenting citizens in Denmark with a patient *dissatisfaction* measure led to more negative views of public hospitals, in contrast to exposing them to a logically equivalent *satisfaction* metric. The purpose of this article is to examine whether public servants are prone to the same kind of bias. Thus, this article aims at answering the following research question: *Are experienced public managers and professionals, when evaluating performance information, susceptible to framing effects?* Knowledgeable public policy and management professionals may be less prone to framing effects because they are likely to be much more familiar with performance metrics than the general public, and also more experienced at making rational decisions based on quantitative information. To address this question, we conducted an experiment with 191 public service professionals in the U.S. in which we experimentally varied the framing of performance information about customer satisfaction, job satisfaction and goal achievement for various federal government agencies.

The next section discusses the theoretical underpinnings of equivalence framing and the judgment of performance information. Next, the experimental design and the participants of the study are characterized, followed by the analysis and presentation of results. Finally, we conclude with a discussion of our findings and their implications for understanding the use of performance information in public administration.

Equivalence framing and managerial assessment of performance

While performance information is celebrated as an "objective," "hard," or "unambiguous" way of communicating the efforts and accomplishments of government, there are considerable degrees of freedom in how exactly to package and show this information. Often it is assumed that these superficial differences in the mere presentation of data have little consequence. However, a core insight from psychology has been that framing exerts powerful effects on how individuals evaluate information and in turn form judgments or make substantive decisions (Baumeister et al. 2001; Tversky and Kahneman 1981; Rozin and Royzman 2001). Although there are various types of framing that affect decision making, one of the most basic and influential is equivalence framing.

Equivalence framing denotes the fact that we can construct "objectively equivalent descriptions of the same problem" (Levin, Schneider, and Gaeth 1998:150). The simplest way of varying the description, while holding constant the informational content, is by changing the valence of the information. Specifically, valence in this case refers to whether the information is presented with a positive or negative frame. Valence is fundamental to how individuals process information. Thus, we would expect that an equivalence framing effect will occur when "two logically equivalent (but not transparently equivalent) statements of a problem lead decision makers to choose different options" (Rabin 1998:36).

Equivalence framing effects are driven by the initial encoding of the information by the recipient, which is affected by the valence of the information. The encoding in turn affects the

cognitive associations that are most likely to become available to the recipient (Levin and Gaeth 1988). That is, for negative valence labels, more negative associations become available in a person's thought processes; and for positive valence descriptions, more positive associations are recalled. In the end, these associations are reflected in the overall assessments, judgments or decisions made by the person based on the information (Levin et al. 1998).

As a response to the plethora of operational definitions of valence framing effects, Levin et al. (1998) propose a taxonomy and distinguish three types, which mainly differ on what is framed, the reaction to the frame, and how such consequences are examined. First, they identify the *risky choice framing*, where subjects are presented with a set of options varying in their risk levels. Usually a hypothetical scenario is exhibited, and two alternatives are given for each frame: one is the sure prospect, while the other implies an all-or-none risky option with clear probabilities assigned. Then, the risk inclination of people is estimated by comparing the choices for risky prospects. Likewise, the *goal framing* looks at the effects on persuasion of people and frames the implication or suggested goal of a behavior to be adopted. This form generally shows subjects the benefits of taking a specific action (positive frame) or exposing them to the loss of not performing it (negative frame). In this case, the persuasive power of the messages is measured by contrasting the rates of adoption. Finally, the *attribute framing* considers how information processing is affected by the descriptive valence of objects or events' characteristics. People are typically presented with frames containing evaluation data, either a percentage of success or the complementary percentage of failure. Subsequently, they assess the object or event trough scales of favorability or yes/no judgements, and the biases produced are estimated by comparing these assessments. Differently from the risky choice framing and the goal framing, this latter type does not stress personal or collective consequences in the frames.

Thus, when evaluating performance information, we are most often concerned with how the information shapes managerial assessments of or judgments about an organization or a unit. From an equivalence framing perceptive, we can think of this as an attribute framing. As such, we manipulate the descriptive valence of performance information in order to capture how it is reflected in the overall evaluation of an organization given the data at hand (Levin et al. 1998). In short, we would expect that the valence of the performance information will lead to a valence consistent shift in the evaluative judgments about the organization and its accomplishments.

For instance, Olsen (2015a) relies on attribute equivalence framing by randomly assigning citizens to either a patient *satisfaction* or a *dissatisfaction* rate, which are logically equivalent; and then asking citizens to evaluate a hospital given this information. He finds that the two-opposing framings produce very different evaluations for an equivalent level of (dis)satisfaction. In accordance with his expectations, the negative label of dissatisfaction provides a negative rating of the hospital, whereas the positive label of satisfaction produces a much more positive rating. The average difference in ratings due to framing is similar to a 20-point change in the underlying citizens' (dis)satisfaction score. The finding of a strong valence framing effect has been replicated in other large samples of nationally representative citizens and with other types of performance data (Olsen 2015b).

The work by Olsen (2015a, 2015b) focused on ordinary citizens, who presumably are not regular or sophisticated consumers of government performance information. Thus, the main question of this study is: *Are public managers, who generally have specialized experience interpreting performance information for decision-making purposes, susceptible to the attribute equivalence framing effect?* We should note that in the context we define public managers broadly, including professionals in both the public and nonprofit sectors as well as other professionals involved in the increasingly complex networks responsible for providing public services in most advanced, post-industrial democracies.

On one hand, we might expect that public managers (being human) would generally remain susceptible to the same cognitive heuristics and biases as citizens. On the other hand, due to their

professional expertise, we might expect public managers to be more reflective and counter such biases when using government performance information to make decisions. Indeed, Moynihan (2008) acknowledges such expertise as one of the basic premises of the performance management movement, particularly in the context of New Public Management (NPM). NPM advocates for business-type managerialism in the public sector, and therefore for increasing the discretion of managers through a reduction of their limitations to reorganize human and fiscal resources from traditional bureaucratic systems (Hood 1991; Osborne and Gaebler 1992). Furthermore, public professionals and managers regularly receive specialized training on performance management and are continuously exposed to interpreting performance metrics and making rational decisions based on this information (Holzer and Yang 2004; Behn 2005).

Research by public management scholars has confirmed the relevance of cognitive and behavioral factors when interpreting performance information. For example, Kroll (2015) shows that cognitive and behavioral factors are relevant for performance information use by public managers and for the organizational circumstances affecting decision making. Moynihan and Pandey (2010) employ survey data from municipal managers in the U.S. to confirm the importance of public service motivation rather than extrinsic motivations in driving the use of performance data. A study from Italy by Bellé, Cantarelli, and Belardinelli (2018) finds that public servants are susceptible to framing effects as well as to other cognitive biases. And in another experimental study of Italian municipal officials, evidence suggests that framing effects may be higher in ex-ante uses of information than in ex-post uses of information (Belardinelli et al. 2018). Our study builds on this evidence of cognitive biases and framing effects in the decision-making of public service professionals and extends it to a U.S. context and to the interpretation of realistic performance indicators for actual American federal agencies. Thus, based on the available evidence, we offer the following hypothesis:

H: Public service professionals and managers are susceptible to equivalence framing effects

Although Olsen (2015a) confirms a strong valence framing effect on citizens when judging performance metrics, he found little evidence that the underlying level of performance influenced the magnitude on such biases. In other words, differences in the numerical rates of (dis)satisfaction shown to citizens under the logically equivalent positive or negative valences did not intensify or diminish their framing effect. Although we have no clear expectation of how this may differ (or not) for public managers, we nevertheless explore how varying the level of performance influences the equivalence framing biases of public managers and professionals when they interpret performance information.

Experimental design and participants

To probe this hypothesis, we designed an online survey experiment that both replicates and extends Olsen's (2015a) study but applied to a sample of U.S. public service professionals ($n = 198$). Given the large magnitude of effects found in previous research, we expect that this sample size should be sufficient for identifying potential framing effects. As shown in Table 1, we presented participants with three vignettes describing different performance indicators for U.S. federal agencies. Specifically, the vignettes referred to customer (dis)satisfaction in the Internal Revenue Service (IRS), job (dis)satisfaction in the Environmental Protection Agency (EPA), and performance targets (not) met in the Social Security Administration (SSA). Although at different levels, in general these measures are familiar to U.S. public managers and related professionals who work with these or similar agencies. For instance, since 2002 federal employees are regularly asked about their job satisfaction on the annual *Federal Employee Viewpoint Survey* conducted by the Office of Personnel Management (U.S. Office of Personnel Management [OPM] 2016). Also, federal, state and local public agencies systematically prepare annual performance plans and

Table 1. Experimental design and vignettes.

Framing	Vignette wording	Percentage
Positive ($n_p = 87$)	The American Customer Satisfaction Index (ACSI) conducts an ongoing national survey of citizen satisfaction with various government agencies. Suppose the latest results for the Internal Revenue Service (IRS) show that X_p percent of its customers are satisfied. *How would you rate the performance of the IRS in this area?*	$X_p \in U[75,\ 95]$
	The Office of Personnel Management (OPM) conducts an annual survey of all federal government workers. Suppose the latest results for the Environmental Protection Agency (EPA) show that X_p percent of its employees are satisfied with their job. *How would you rate the performance of the EPA in this area?*	$X_p \in U[75,\ 95]$
	Federal agencies are required to report progress on their performance goals and targets each year in their Performance and Accountability Reports (PARs). Suppose the latest report on the Social Security Administration (SSA) shows that X_p percent of its performance targets were met last year. *How would you rate the performance of the SSA in this area?*	$X_p \in U[75,\ 95]$
Negative ($n_n = 103$)	The American Customer Satisfaction Index (ACSI) conducts an ongoing national survey of citizen satisfaction with various government agencies. Suppose the latest results for the Internal Revenue Service (IRS) show that X_n percent of its customers are dissatisfied. *How would you rate the performance of the IRS in this area?*	$X_n \in U[5,\ 25]$
	The Office of Personnel Management (OPM) conducts an annual survey of all federal government workers. Suppose the latest results for the Environmental Protection Agency (EPA) show that X_n percent of its employees are dissatisfied with their job. *How would you rate the performance of the EPA in this area?*	$X_n \in U[5,\ 25]$
	Federal agencies are required to report progress on their performance goals and targets each year in their Performance and Accountability Reports (PARs). Suppose the latest report on the Social Security Administration (SSA) shows that X_n percent of its performance targets were not met last year. *How would you rate the performance of the SSA in this area?*	$X_n \in U[5,\ 25]$

Note: The column "Percentage" shows the ranges of the uniform random variables for the negative frame [X_n] and the positive frame [X_p].

reports that contain specific performance targets and their accomplishment (Melkers and Willoughby 1998; Latham, Borgogni, and Petitta 2008; Poister and Streib 2005; Joyce 2011). Likewise, the concept of customer satisfaction and its measurement have become increasingly important across government agencies, especially at the federal level after the *National Performance Review* initiative in the early 1990s (Fountain 2001). Finally, we used specific and well-known U.S. public agencies for each vignette, trying to resemble a real-world managerial context.

Respondents were randomly assigned to either a positive or negative frame for all three-performance metrics chosen for this study (again, see Table 1). For example, for the SSA, some participants were told that the latest results showed that "83% of its performance targets were met last year," while others were told that "17% of its performance targets were not met last year." The percent of targets meet, we should note, is a key feature of the federal government's Performance and Accountability Reports (PAR) program. The values of the indicators (percentages) shown were randomized, as described shortly. After reading each vignette, respondents were asked to rate the performance of each agency–in the specific area the information exhibited was about–using a horizontal slider ranging from 0 (very bad) to 100 (very good).

We should note that we considered including a third "control" arm of the experiment that would be presented with both the positive and negative frames for all metrics (e.g., "83% of SSA's performance targets were met last year, while 17% of its performance targets were not met"), but for reasons of sample size decided against it. Moreover, this additional arm would have implied its own complications, such as the word-order of the performance data shown (whether

presenting the positive and then the negative frame, or vice versa, or even randomizing the order). As a result, we decided to just rely on the positive and negative framing conditions, as was done in Olsen's (2015a) original study as well as many other studies of framing effects (Chong and Druckman 2007).

Three randomized elements were used for the experiment. First, the positive or negative framing was randomly assigned (at the respondent level). Hence, some participants received positively framed information for the EPA, the SSA and the IRS ($n_p = 87$), while others received negatively framed information ($n_n = 103$). In addition, we utilized a random variable X_i ($i = n, p$) to represent various percentages for the performance metrics in each vignette. X_i follows a uniform distribution, with every value in the interval being equally likely to occur. As shown in Table 1, the random variable for the negative frame (X_n) ranged between 5 to 25%, and the one for the positive frame (X_p) fluctuated between 75 and 95%. Finally, the order of the three vignettes was also randomized.

To be sure, based on the taxonomy proposed by Levin et al. (1998), the three vignettes of this experiment correspond to the features of attribute framing effects. First, the framing impact on risk preferences of public service professionals is not a concern here, neither the effect on the persuasive power of messages for their adoption of certain behavior. Thus, the frames in the vignettes did not stress individual or collective implications. Rather, in all of them the descriptive valence of performance information representing specific managerial areas of public agencies were manipulated by showing percentages of success in the positive frames (i.e., satisfaction, targets met), or their complementary percentages of failure in the negative frames (i.e., dissatisfaction, targets not met). Moreover, subjects were thereafter requested to rate the agencies in such managerial areas.

Participants were respondents to an emailed survey sent to members of the *Public Service Research Panel (PSRPanel)*, a university-affiliated online panel (see PSRPanel.org). The PSRPanel continuously recruits public managers and related professionals working for the public and the nonprofit sectors via professional networks such as mailing lists of relevant professional associations, social media, and advertising at professional conferences, as well as recruiting from graduate programs that specifically train public and nonprofit managers. The Panel includes mainly U.S. public service professionals of diverse sociodemographic characteristics and from various geographic regions.

We implemented the survey experiment in two waves about a year apart, largely because we were growing the panel at the time and wanted to include new panelists to have more statistical power. First, we sent 572 email invitations to all existing members of the PSRPanel during May 2015, requesting them to answer a survey about judgment of performance information and a few related topics in public service. Then, a year later, we sent an additional 69 email invitations to new panel members. A total of 198 panelists responded to the survey, representing a 31% response rate. Since the vignettes focused on U.S. government agencies, 7 non-U.S. respondents were excluded from the analytical sample. Descriptive statistics for the resulting 191 eligible respondents are shown in Table 2. In general, participants are predominantly females (59%) and white (72%) with most falling into the 40–59 age category (55%). The majority have a graduate degree (77%), work for the public and nonprofit sectors (80%) and hold a managerial position (75%). The 25% of subjects who are not managers or supervisors are either graduate students (in training for public and nonprofit managerial jobs), people in technical jobs (e.g., specialists or analysts), self-employed (consultants), or retired (from professional jobs). Similarly, the 20% who do not work for the public or nonprofit sectors are mostly professionals who work for consulting firms, universities, or are self-employed. We decided to keep this diverse group of professionals in our sample for statistical reasons but also because of our broad definition of public managers as including public, nonprofit, and even private-sector professionals involved in the modern provision of public services.

Table 2. Descriptive statistics.

Variable	Percentage	n
Gender		
Male	40.8%	62
Female	59.2%	90
Age		
20-29	10.6%	16
30-39	17.9%	27
40-49	27.2%	41
50-59	27.8%	42
60-69	14.6%	22
70 or older	2%	3
Race		
White	71.9%	110
Black	15.7%	24
Hispanic	5.9%	9
Asian	2.6%	4
Other	3.9%	6
Education		
Graduate degree	76.6%	118
College	22.1%	34
Less than College	1.3%	2
Position		
Top manager	22.8%	29
Middle manager	32.3%	41
Team leader	12.6%	16
Supervisor	7.1%	9
None (not a manager or supervisor)	25.2%	32
Employment		
Public sector	61.4%	97
Non-profit sector	18.4%	29
Private (for-profit sector)	3.8%	6
Independent consultant or self-employed	3.2%	5
Student/retired/unemployed	11.4%	18
Other	1.9%	3
Geographic location		
Northeast	39.9%	61
South	23.5%	36
Midwest	16.3%	25
West	20.3%	31

Analysis and results

Our analytical approach will proceed as follows. By using kernel density estimates, we will first examine the distribution of responses and contrast it in the positive and negative frames for each vignette. We then conducted t-tests for each vignette comparing the means of performance ratings for both frames. Finally, we used a set of regression models for each vignette: one includes just the framing variable assigned to public servants (*Model 1*), while the others add the percentage level shown to them (*Model 2*), and the interaction between these two variables (*Model 3*). All the analyses were conducted using Stata 14.

The density plots in Figures 1–3 represent the distributions of performance judgments made by participants. The x-axis depicts their evaluative judgments, from 0 (very bad) to 100 (very good). In Figure 1, showing the results for the IRS and customer satisfaction, the pattern suggests a greater density of high ratings when the agency's performance is framed in terms of the percentage of *satisfied* customers (the positive frame), compared with a logically equivalent percentage of *dissatisfied* customers (the negative frame). Similarly, in Figure 2 for the EPA, the positive framing in terms of *job satisfaction* results in higher assessments of the agencies performance, relative to the logically equivalent level of *job dissatisfaction*. And again, in Figure 3, the framing

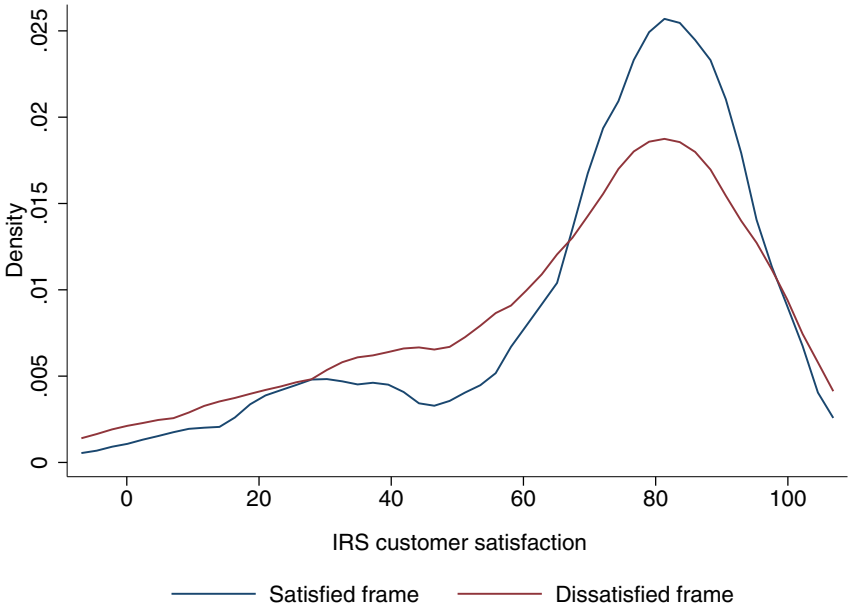

Figure 1. Distribution of performance ratings for IRS customer satisfaction (kernel density estimates).

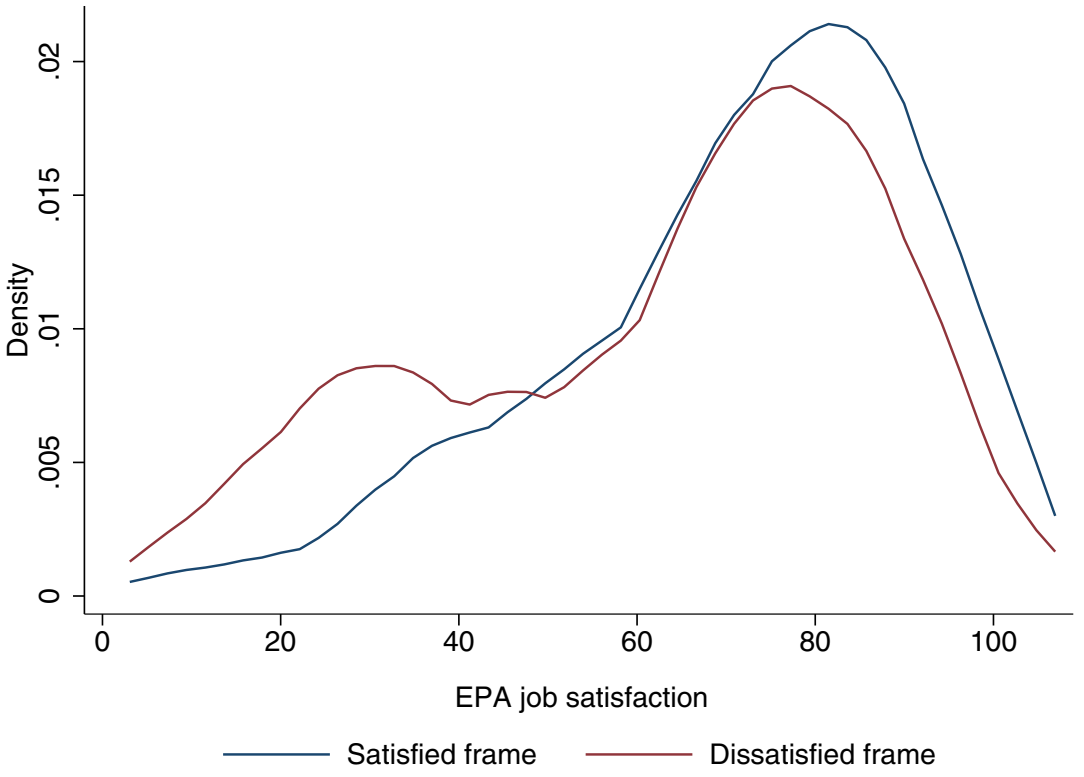

Figure 2. Distribution of performance ratings for EPA job satisfaction (kernel density estimates).

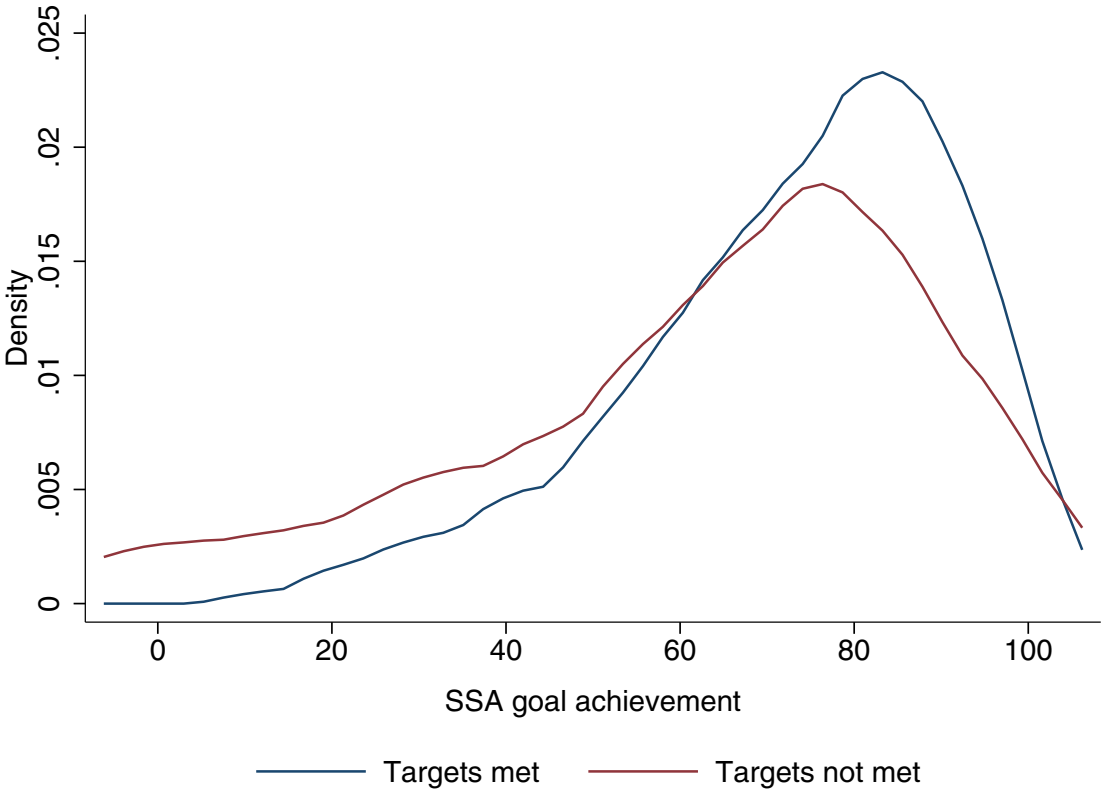

Figure 3. Distribution of performance ratings for SSA targets met (kernel density estimates). *Note*: The area under the curves in the density plots of Figures 1–3 represents the percentage of the data between certain ranges of performance ratings from respondents on the *x*-axis.

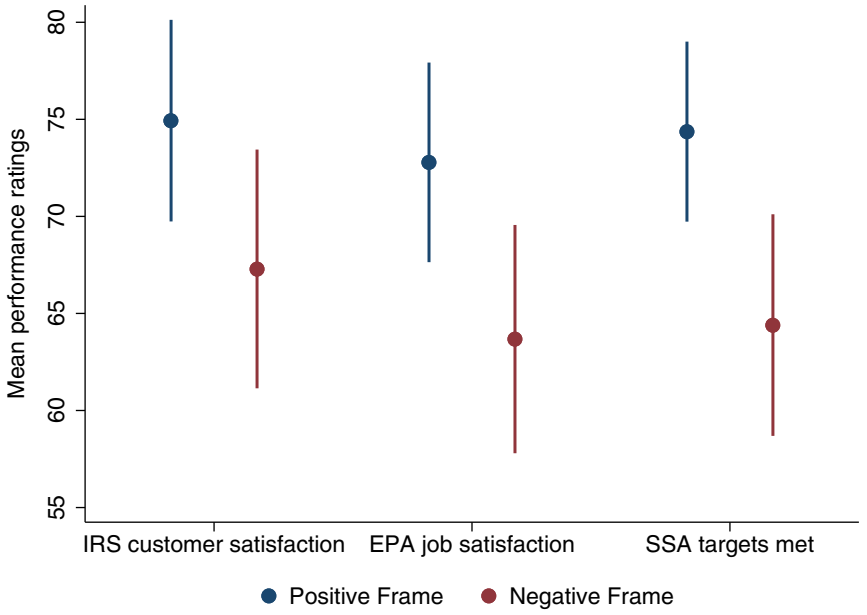

Figure 4. Difference in mean ratings for each performance indicator with 95% confidence intervals. *Note:* The difference in means for job satisfaction was significant at 0.05, and the difference in means for goal achievement was significant at 0.01.

of SSA's performance in terms of *targets met* (the positive frame) leads respondents to rate the agency as doing better, compared with the logically equivalent information presented in terms of *targets not met* (the negative frame). These three graphs clearly show a framing effect in how public service professionals judge the performance of these three federal agencies.

Figure 4 shows the differences in means obtained for each performance rating—by public service managers and professionals—comparing the positive and negative frames. The average for goal achievement is 67.8, for job satisfaction is 67.0, and for customer satisfaction is 68.6. The mean assessment for goal achievement is 73.2 when participants are exposed to positively framed performance information (*percentage of targets met*), but only 62.8 when they are presented with logically equivalent performance metrics that are negatively framed (*percentage of targets not met*). This represents a statistically significant difference of 10.4 points between both frames ($p = 0.006 < 0.01$). Similarly, the mean assessment for job satisfaction is 71.5 when participants are shown positively framed performance information (*percentage of job satisfaction*), but only 62.7 when they are exposed to the corresponding performance data that is negatively phrased (*percentage of job dissatisfaction*). This difference of 8.9 points is also significant ($p = 0.017 < 0.05$). For customer satisfaction, the mean assessment of 71.3 points under the positive frame (*percentage of customer satisfaction*) and 66.0 points under the negative frame (*percentage of customer dissatisfaction*) are in the expected direction and follow the pattern of the other two performance measures, although the difference in means is not statistically significant ($p = 0.189$).

To further test the statistical significance of these patterns, Table 3 shows a set of regression models for each vignette. *Model 1* includes just the main effect of the framing ($1 =$ negative frame; $0 =$ positive frame). *Model 2* adds the level of the performance metric shown to respondents, that is the level of customer satisfaction (for the IRS), job satisfaction (for the EPA), and targets met (for the SSA). It should be noted the value of this variable was coded consistently in the positive direction; for example, the 10% level of "dissatisfied" customers in the IRS vignettes was coded as 90 (its equivalent level had it been positively framed). *Model 3* considers the interaction of the framing and the level of the performance metric, and thus captures whether any framing effect varies with the underlying level of performance.

The first set of regressions in Table 3 are for the IRS and customer satisfaction, and they show a framing effect in the expected direction, but not statistically significant.[1] Specifically, as model 1 shows, the framing of performance information in terms of dissatisfaction produces an average decline of about five points in the performance evaluations of the IRS given by respondents. The treatment effect is slightly larger in models 2 and 3 and fairly close to being statistically significant. As expected, the level has a positive effect, meaning that respondents gave higher (lower) evaluations to vignettes showing more satisfaction (less dissatisfaction) by IRS customers. There is no evidence of an interaction effect of the treatment (framing) and the level of satisfaction.

The next set of regressions in Table 3 refer to the EPA and job satisfaction. In all three models, the impact of the framing is negative and significant. Indeed, when respondents are exposed to the dissatisfaction frame instead of the corresponding satisfaction frame, their evaluations of the EPA decrease on average by nearly nine points. This framing effect is statistically significant in all three models, and also substantively large (a 13% decline) relative to the constant (which in Model 1 is the average rating of the EPA's performance under the positive job satisfaction framing). Turning to the level, the effect is again positive–as expected–and significant in both models 1 and 2. However, the interaction of the framing and the level is close to zero and nonsignificant.

The final set of regressions in Table 3 are for the percentage of targets met by the SSA according to its Performance and Accountability Report. When this metric is framed as goals not met, contrasted with logically equivalent goals met, respondents evaluated the SSA 10–11 points lower on the 0–100 performance rating scale. This negative framing effect is highly significant statistically and also substantively large relative to the constant, representing about a 14% decline on average in the performance evaluations. The level effect is again positive and significant, which

Table 3. Regression results.

	IRS customer satisfaction			EPA job satisfaction			SSA targets met		
	Model 1	*Model 2*	*Model 3*	*Model 1*	*Model 2*	*Model 3*	*Model 1*	*Model 2*	*Model 3*
Framing (1 = negative)	−5.345	−6.167	−6.206	−8.851**	−8.746**	−8.783**	−10.413***	−11.028***	−10.853***
	(4.055)	(4.014)	(4.029)	(3.676)	(3.562)	(3.575)	(3.746)	(3.495)	(3.508)
Level		0.78**	0.678		0.982***	1.089**		1.337***	1.087**
		(0.341)	(0.495)		(0.309)	(0.45)		(0.286)	(0.433)
Framing* Level			0.195			−0.202			0.445 (0.577)
			(0.685)			(0.62)			
Constant	71.319***	71.559***	71.528***	71.522***	71.651***	71.664***	73.224***	74.021***	73.872***
	(2.915)	(2.876)	(2.887)	(2.636)	(2.554)	(2.562)	(2.714)	(2.536)	(2.547)
Adjusted R	0.005	0.033	0.027	0.033	0.092	0.086	0.046	0.171	0.168
F-statistic	1.74	3.51	2.35	5.8	8.15	5.44	7.73	15.4	10.43
N		149			142			141	

*$p < 0.1$, **$p < 0.05$,***$p < 0.01$.

Note: Standard errors are in parentheses. *Model 1* considers just the frame assigned to respondents; *Model 2* adds the percentage level shown to them; and *Model 3* also includes the interaction between the framing variable and the percentage level.

suggests that when more targets are being met (or fewer unmet), the SSA is rated as performing better (or worse). As in with the other vignettes, the effect of the interaction term is negligible and not significant.

Discussion and implications

Our study contributes to a growing body of evidence that framing effects matter in the interpretation of government performance information (Boyne et al. 2009; James 2011; James and Moseley 2014). Building on work by Olsen (2015a) with citizens, we find similar framing effects shape the evaluative judgments of presumably more sophisticated and experienced professional public managers. Thus, our study helps confirm recent emerging evidence about how cognitive biases and framing effects, in particular, shape how public managers interpret and use performance information (Bellé et al. 2018; Belardinelli et al. 2018). More broadly, our findings support the growing understanding of how cognitive factors affect the uses of performance information by decision makers in public organizations (Moynihan and Pandey 2010; Kroll 2014, 2015). But before discussing our results and their implications for public management, it is important to acknowledge several key limitations of our study.

To begin with, we relied on an online panel of public managers and related professionals who voluntarily signed up and therefore are self-selected. They were mainly recruited from relevant professional networks and from public affairs graduate programs, and therefore they are likely to have higher research interests and better analytical training than a broader cross-section of public workers. In addition, although we tried to use authentic and widely known U.S. agencies and performance metrics, for practical reasons we did not adapt the experiment to each person's actual work context. Thus, naturally, the tasks requested in the vignettes lacked meaningful consequences and realism for participants. This raises the external validity concern about how our findings translate into participants' actual work settings and decision making. Likewise, it is possible that some indicators we used, especially customer satisfaction and job satisfaction, may not have been clearly interpreted as dichotomous. Specifically, 75% satisfaction may not necessarily signify an equivalent 25% dissatisfaction if respondents imagined a middle category like "neither satisfied nor dissatisfied." But our experimental design was modeled on Olsen's (2015a) study, and indeed one of our main motivations was to replicate and extend this prior research. Moreover, the strongest framing bias found was for targets met, versus targets not met, which is clearly a more dichotomous performance metric.

Overall, our findings support our hypothesis and are consistent with the equivalence framing effect, which is well documented in psychology (Levin et al. 1998) and more recently in public administration research involving citizens (Olsen 2015a) and public service professionals (Bellé et al. 2018; Belardinelli et al. 2018). We add to this body of research by testing experienced public service managers and professionals and find that, like regular citizens, they are prone to being influenced by superficial changes in the valence of performance information when asked to read and interpret it. More precisely, results from two out of three metrics used in our experiment confirm equivalence framing effects (targets met and job satisfaction). The results for customer satisfaction are in the expected direction, although they are not as strong as the other two vignettes performance and therefore marginally insignificant. These non-significant results would probably be significant for a larger sample size.

Moreover, customer satisfaction is a metric directly relevant for respondents' own experiences, as most of them are undoubtedly taxpayers who themselves may be satisfied or dissatisfied with the Internal Revenue Service (IRS). Thus, they might have rated the performance of the IRS partially based on their own interactions with this agency or their feelings about it, and not just the information offered in the vignette. Furthermore, people tend to have strong (often partisan) feelings about the IRS, which tends to receive low favorability ratings in public opinion surveys.[2] In short, for the

customer satisfaction metric, respondents' personal views about the IRS may have had a larger influence on their performance judgments and thus attenuated the framing effect.

Taken together, our results suggest that the negative framing of performance information of well-known federal agencies lead public service professionals to judge performance less favorably, while positive framing leads them to form more favorable evaluations of these same organizations.[3]

The main framing effects in our study, however, are only about half the size as found among citizens in Olsen's (2015a) study. This finding is consistent with the hypothesis that public managers may be at least somewhat more reflective and less susceptible to framing effects. Moreover, this result is consistent with Olsen's (2015a) finding that the framing effect on citizens in his study was reduced by their current or previous work experience in public organizations. This potential reduction in bias from equivalence framing may be attributable to the training of public managers, as well as perhaps their professional experience with using performance information. In an attempt to probe for an explanation, we tried various interactions between the treatment (framing) and several background variables, such as years of experience, position, and employment (see Table 2). We also tried interactions with self-perceptions of expertise in several management and technical areas (such as strategic planning, performance management and data analysis) as proxies of their familiarity with the performance indicators used in the vignette. The results for these interaction tests in general were not consistent and reliable—especially considering the sample size of our study—and thus we do not present them here. Certainly, this topic deserves more research and may lead to a better understanding of what types of training or professional experiences lead public managers to overcome framing effects when interpreting performance metrics, at least in part.

In any case, the results highlight the value of explicitly unifying the study of how two types of end-users of performance information, namely citizens and public service managers and professionals, process these metrics. Their motivation for engaging with performance data might be very different, as—for instance—framing can be used or even abused to influence decision making. In any event, they still share a common human cognition that helps them to process the information. Moreover, by explicitly comparing citizens and managers, we can learn some very fundamental principles of performance information use and effects. These will be stressed as we find strong similarities in how very different users process information. At the same time, any differences in effects are worth studying to find the underlying reasons. It might be motivations, professionalism, institutional roles, incentives, or some other factors that set citizens and managers apart in the way that performance information affects them.

These findings may be helpful in the design and implementation of performance management systems. One main agreement on the performance management literature is the huge challenge its effective adoption poses for public agencies (Behn 2002; Moynihan 2008). Hence, once performance information is actually used by public managers and professionals as well as by other stakeholders, the next step should be to prevent them from unintended consequences of framing so that they engage in better decision-making. This study might help in this purpose since the findings suggest possible interesting venues of research aimed at de-biasing or minimizing framing effects. For instance, we might ask if certain types of performance information can help reveal potential framing effects before managers interpret or act on the information.

Notes

1. The p-values for customer satisfaction are $p = 0.189$ (model 1), $p = 0.127$ (model 2) and $p = 0.126$ (model 3).
2. For instance, survey results by the Pew Research Center in 2019 reflected this unfavourability of the IRS—along with the Immigration and Customs Enforcement (ICE), the Department of Education, the Department of Veteran Affairs and the Department of Justice—compared to other federal agencies such as

the U.S. Postal Service, the National Park Service, the National Aeronautics and Space Administration (NASA) and the Centers for Disease Control and Prevention (CDC). For further information, please visit https://www.people-press.org/2019/10/01/public-expresses-favorable-views-of-a-number-of-federal-agencies/.

3. Although more specific subgroup analyses are difficult due to sample size limitations, as a robustness check, we examined these results for the sub-group of people who specifically declared to work in the federal, state, or local government. The results hold for this sub-group as the equivalence framing effect is statistically significant for all performance metrics used in the experiment. They are even slightly higher than the effects obtained for the broader sample of public servants.

Disclosure statement

The authors declare that they have no relevant or material financial interests that relate to the research described in this article.

Funding

Javier Fuenzalida would like to thank the funding received for this article by the Center for Research in Inclusive Education (PIA-Conicyt CIE 160009).

Notes on contributors

Javier Fuenzalida (jafuenza@dii.uchile.cl) is assistant professor in the Institute of Public Affairs at the University of Chile. He is also research fellow of the Centre for Public Systems and the Center for Research in Inclusive Education (PIA-CONICYT CIE160009), both at the Industrial Engineering Department at the University of Chile. His research interests are in public management, civil service and performance management.

Gregg G. Van Ryzin (vanryzin@rutgers.edu) is professor in the School of Public Affairs and Administration at Rutgers University in Newark. His research focuses on the use of surveys and experimental methods to understand how citizens evaluate and respond to public services and institutions.

Asmus Leth Olsen (ajlo@ifs.ku.dk) is professor in the Department of Political Science at the University of Copenhagen. His research focuses on the effects of performance information, political and administrative psychology, behavioral public administration, and the application of experimental methods in public administration.

References

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.

Baumeister, R. F., E. Bratslavsky, C. Finkenauer, and K. D. Vohs. 2001. "Bad is Stronger than Good." *Review of General Psychology* 5(4):323–70. doi: 10.1037/1089-2680.5.4.323.

Behn, R. D. 2002. "The Psychological Barriers to Performance Management: Or Why Isn't Everyone Jumping on the Performance-Management Bandwagon?" *Public Performance & Management Review* 26(1):5–25. doi: 10.1080/15309576.2002.11643684.

Behn, R. D. 2005. "The Core Drivers of CitiStat: It's Not Just about the Meetings and the Maps." *International Public Management Journal* 8(3):295–319. doi: 10.1080/10967490500439594.

Belardinelli, P., N. Bellé, M. Sicilia, and I. Steccolini. 2018. "Framing Effects under Different Uses of Performance Information: An Experimental Study on Public Managers." *Public Administration Review* 78(6):841–51. doi: 10.1111/puar.12969.

Bellé, N., P. Cantarelli, and P. Belardinelli. 2018. "Prospect Theory Goes Public: Experimental Evidence on Cognitive Biases in Public Policy and Management Decisions." *Public Administration Review* 78(6):828–40. doi: 10.1111/puar.12960.

Boyne, G. 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):1273–84. doi: 10.1017/S0022381609990089.

Chong, D., and J. N. Druckman. 2007. "Framing Theory." *Annual Review of Political Science* 10(1):103–26. doi: 10.1146/annurev.polisci.10.072805.103054.

Fountain, J. E. 2001. "Paradoxes of Public Sector Customer Service." *Governance* 14(1):55–73. doi: 10.1111/0952-1895.00151.

Hatry, H. P. 2006. *Performance Measurement: Getting Results*. Washington, DC: The Urban Institute.

Holmes, M., and D. Shand. 1995. "Management Reform: Some Practitioner Perspectives on the past Ten Years." *Governance* 8(4):551–78. doi: 10.1111/j.1468-0491.1995.tb00227.x.

Holzer, M., and K. Yang. 2004. "Performance Measurement and Improvement: An Assessment of the State of the Art." *International Review of Administrative Sciences* 70(1):15–31. doi: 10.1177/0020852304041228.

Hood, C. 1991. "A Public Management for All Seasons?" *Public Administration* 69(1):3–19. doi: 10.1111/j.1467-9299.1991.tb00779.x.

James, O. 2011. "Managing Citizens' Expectations of Public Service Performance: Evidence from Observation and Experimentation in Local Government." *Public Administration* 89(4):1419–35. 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.

Joyce, P. G. 2011. "The Obama Administration and PBB: Building on the Legacy of Federal Performance-Informed Budgeting?" *Public Administration Review* 71(3):356–67. doi: 10.1111/j.1540-6210.2011.02355.x.

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.

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. 2015. "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.

Kroll, A., and D. Vogel. 2014. "The PSM–Leadership Fit: A Model of Performance Information Use." *Public Administration* 92(4):974–91. doi: 10.1111/padm.12014.

Latham, G. P., L. Borgogni, and L. Petitta. 2008. "Goal Setting and Performance Management in the Public Sector." *International Public Management Journal* 11(4):385–403. doi: 10.1080/10967490802491087.

Levin, I. P., and G. J. Gaeth. 1988. "How Consumers Are Affected by the Framing of Attribute Information before and after Consuming the Product." *Journal of Consumer Research* 15(3):374–8. doi: 10.1086/209174.

Levin, I. P., S. L. Schneider, and G. J. Gaeth. 1998. "All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects." *Organizational Behavior and Human Decision Processes* 76(2):149–88. doi: 10.1006/obhd.1998.2804.

Melkers, J., and K. Willoughby. 1998. "The State of the States: Performance-Based Budgeting Requirements in 47 out of 50." *Public Administration Review* 58(1):66–73. doi: 10.2307/976891.

Moynihan, D. P. 2008. *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, DC: Georgetown 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. 2012. "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.

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. "Negative Performance Information Causes Asymmetrical Evaluations and Elicits Strong Responsibility Attributions." Paper presented at the 111th Annual Meeting of the American Political Science Association, September 3–6, San Francisco, CA.

Osborne, D., and T. Gaebler. 1992. *Reinventing Government: How the Entrepreneurial Spirit is Transforming Government*. Reading, MA: Addison Wesley Public Comp.

Poister, T. H., M. P. Aristigueta, and J. L. Hall. 2014. *Managing and Measuring Performance in Public and Nonprofit Organizations: An Integrated Approach*. 2nd ed. San Francisco, CA: John Wiley & Sons.

Poister, T. H., and G. Streib. 2005. "Elements of Strategic Planning and Management in Municipal Government: Status after Two Decades." *Public Administration Review* 65(1):45–56. doi: 10.1111/j.1540-6210.2005.00429.x.

Rabin, M. 1998. "Psychology and Economics." *Journal of Economic Literature* 51(2):528–46. doi: 10.1257/jel.51.2.528.

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.

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.

Taylor, J. 2011. "Factors Influencing the Use of Performance Information for Decision Making in Australian State Agencies." *Public Administration* 89(4):1316–34. doi: 10.1111/j.1467-9299.2011.02008.x.
Tversky, A., and D. Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211(4481):453–8. doi: 10.1126/science.7455683.
U.S. Office of Personnel Management [OPM]. 2016. *Federal Employee Viewpoint Survey* (Technical Report). Washington, DC: U.S. Office of Personnel Management.
Van Dooren, W., G. Bouckaert, and J. Halligan. 2015. *Performance Management in the Public Sector*. 2nd ed. New York, NY: Routledge.
Van Dooren, W., and S. Van de Walle (Eds.). 2008. *Performance Information in the Public Sector: How It Is Used*. New York, NY: Palgrave Macmillan.