



## **Public Sector Communication and Performance Management**

Olsen, Asmus Leth

*Published in:*  
Public Sector Communication

*Publication date:*  
2018

*Document version*  
Early version, also known as pre-print

*Citation for published version (APA):*  
Olsen, A. L. (2018). Public Sector Communication and Performance Management: Drawing Inferences from Public Performance Numbers. In *Public Sector Communication*



Københavns Universitet



## Public Sector Communication and Performance Management

Olsen, Asmus Leth

*Published in:*  
Public Sector Communication

*Publication date:*  
2018

*Document Version*  
Early version, also known as pre-print

*Citation for published version (APA):*  
Olsen, A. L. (2018). Public Sector Communication and Performance Management: Drawing Inferences from Public Performance Numbers. In Public Sector Communication

### **3.1 Public Sector Communication and Performance Management:**

#### **Drawing Inferences from Public Performance Numbers**

Asmus Leth Olsen, Department of Political Science, University of Copenhagen

Øster Farimagsgade 5, opgang E, DK-1353, Copenhagen K, Denmark. Email: [ajlo@ifs.ku.dk](mailto:ajlo@ifs.ku.dk)

*First draft. December 27<sup>th</sup>, 2016.*

*This draft. January 31<sup>st</sup>, 2018*

*Number of Words: 7,389*

**Abstract:** From a public sector communication perspective, performance management introduces a reliance on quantitative performance information intended for both internal and external use. From a communication perspective it raises the question how in particular citizens make sense of performance data? The chapter outlines state of the art on behavioral and experimental research on how citizens' perception of public services is affected by performance information. These studies show the distinct effect of quantitative data on perception, the role of benchmarks and comparisons, the overly reliance on negative information relative to positive, and the importance citizens' own prior beliefs when evaluating performance data. These findings call to attention a need for performance management to rely on psychological insights to understand the effect of performance information when communicating to citizens.

**Keywords:** public management; behavioral public administration; experiments; citizen satisfaction; performance perceptions

## **Introduction**

In the late 1930s a young Herbert Simon was fundamentally curious about how the measurement of the performance of public organizations can be useful to managers and citizens. In an almost forgotten piece from 1939, *The Administrator in Search of Statistics*, he asks a set of simple questions about the role of performance information in the public sector: “*What are statistics and how can they help solve administrative problems? What is the proper use of statistics and what pitfalls are to be avoided?*” (Simon, 1939, pp. 21). While Simon went on to other topics, and later was awarded the Nobel Prize in Economics, his early endeavors into the effects of performance information has been revitalized in recent decades by the performance management doctrine (Moynihan, 2006; Pollitt, 2006). Moynihan (2008, pp. 5) defines performance management as “a system that generates performance information through strategic planning and performance measurement routines and that connects this information to decision venues, where, ideally, the information influences a range of possible decisions.” In performance management performance information is intended to play multiple roles such as a means for evaluating actions, allocate resources, learning, improve decision making and accountability (Behn, 2003, Holzer & Yang, 2004; Askim, 2007). To date meta-analytical evidence have found very little evidence of these endeavors on actual performance (Gerrish, 2015, Hood & Dixon, 2015). However, from a public sector communication perspective, the most important aspect is the external role of performance information, i.e., how information about performance affects citizens’ perception, attitudes, and behavior in relation to public services. As Osborne and Plastrik (2000, pp. 247) put it, performance information should ideally help “citizens and customers judge the value that government creates for them”. However, at this point in the stage of research we are in many ways still occupied with the initial question: What inferences do citizens draw from performance information? Any attempt to constructively communicate performance to the

wider public must rely on some assumptions or actual evidence that helps answer this question.

This chapter focuses on the state of the art of research that aims to understand the external role of performance information. This research emphasizes the need for rigorous studies of how the public sector communicates performance to citizens and how this communication can enhance citizens' understanding of public services. In many ways the aim is to replace assumptions about performance information effects with actual experimental and observation evidence of how citizens rely on performance data when evaluating performance. In fact, at the turn of the 20<sup>th</sup> century Herman C. Beyle in a similar manner called for a communication oriented study of how citizens perceive government performance (Beyle, 1928, Beyle, 1936). However, only recently there has been an emergence of a collective effort to understand the “recipient” of performance information.

Over the recent years, performance information research has marked by a common methodological and theoretical focus. Theoretically, this research draws on insights from psychology and the behavioral sciences (Olsen 2015c, Moynihan 2016, James and Olsen 2017; Andersen and Hjortskov 2016) and therefore responds to Herbert Simon's early calls for a closer collaboration between psychology and public administration (Simon 1955, Olsen 2015a). It exemplifies the recent turn to behavioral public administration which is defined by “the micro-level perspective of individual behavior and attitudes by drawing on recent advances in our understanding of the underlying psychology and behavior of individuals and groups.” (Grimmelikeuisen et al., 2017, p. 46). This was not always the case as James and John (2007, p. 568) where the some of the first to note: “The connection between public management and political behavior has been largely neglected because the literatures have mostly been developed as separate areas of endeavor.” In terms of methods, this literature has become predominately experimental and has thus been one of the leading topics behind the

spread of experimental methods in public administration (Margetts, 2011, Blom-Hansen et al., 2015). Performance measurement research is hereby moving closer to the theoretical and methodological tradition in political communication which has been closely integrated with political psychology. Also, with the rise of “big data” and its increased importance for government affairs all possible questions about communicating public performance becomes increasingly pressing (Lavertu, 2016, Mergel et al., 2016). It calls for a view of the external role of performance information as an interdisciplinary study of communication to citizen which will help foster further theoretical and methodological integration.

In the remainder of this chapter defines the concept of performance information. Then the state of the art of the current research is outlined and some of the major themes. Finally a future agenda for communication research and performance management is discussed with an emphasis on the limits and future potential of this research – in particular the need for theoretical integration and connecting evidence from micro to macro level studies (and back).

### **The Study of Performance Information as a Study of Numbers**

Conceptually “performance information,” “performance indicator,” and “performance measurement” have been used interchangeably in the field which also will be the case in this chapter (Olsen, 2015c). According to Pollitt (2006, p. 39) performance information is defined as “systematic information describing the outputs and outcomes of public programs and organizations whether intended or otherwise generated by systems and processes intended to produce such information.” Van Dooren et al. (2010, p. 6) offers the simple definition that “measuring performance means systematically collecting data.” Along the same lines Martin and Kettner (1997, p. 18) define performance measurement as, “the regular collection and reporting of information on the efficiency, quality, and effectiveness of government

programs” and Wang (2000, p. 104) highlights “the measuring of levels of activities and achievements through a range of indicators.” Many stress that performance information should be a repeated and regular exercise. Hatry (2006, p. 6) sees it as the “regular tracking is a key characteristic of performance measurement”. Thomas (2006, p. 1) defines performance measurement as “the regular generation, collection, analysis, reporting and utilization of a range of data related to the operation of public organizations and public programs, including data on inputs, outputs, and outcomes.” James (2011, p. 400) also includes a broader range of public sector information as he defines performance information as “[...] the inputs, processes, outputs, and outcomes of public services and organizations providing them and their value for money, efficiency, and effectiveness.”

A unifying feature of these definitions and many more is the view of performance information as fundamentally numerical data (Davies, 1999, Hood, 2007, Moynihan, 2008, Perrin, 1998, Van Dooren, Bouckaert, & Halligan 2010, Olsen, 2015c). As Radin (2006, p. 27) notes: “There is perhaps no element within the performance measurement process that is more important than the reliance on numbers and quantitative presentation of accomplishments”. This feature allows us to connect performance measures with the broader debate on quantification in politics (Espeland & Sauder, 2007, Herbst, 1992, Stone, 1997). It has also allowed research to draw on existing studies in other fields that study how humans draw inferences from quantitative data about the social world (Olsen, 2017a). This interdisciplinary turn can be seen in the recent experimental and behavioral performance measure research.

### **State of the Art: Behavioral and Experimental Research**

A natural point of departure is to consider how citizens respond to “hard” statistical performance data compared to others ways obtaining information about performance (Olsen,

2017a). In their everyday lives citizens have numerous casual encounters with public services. They interact with public employees and are exposed to local news on services. Such exposure is often high on vividness and can be emotionally engaging. Hard performance data, on the other hand, is often pallid packages of thematic information and dry statistics – low vividness input which has been found to be less emotionally engaging in other setting (Slovic, 2007, Gross, 2008). This effect has also been noted by communication scholars (Zillerman, 2003). Similarly, exposure to simple descriptions of individual experiences with public services elicit stronger emotional response and shift evaluations of public service more than numerical performance information describing the very same performance level (Olsen 2017a).

These findings actually correspond with some of the first ever evidence of what citizens learn from government performance information. In the early 1940s a graduate student named Miriam Roher ran a number of surveys to understand how citizens are affected by being exposed to performance information via government reports and local news media (Lee 2006). She fielded a number of surveys asking citizens of Palo Alto California about their knowledge on government services and performance (Roher 1941a, Roher 1941b). One of them was fielded prior to the circulation of a multi-page government report which was being distributed to all households. Another survey was fielded just after the reported was distributed. By comparing responses across the two samples she could estimate how the dissemination of statistical performance data affected citizens' knowledge on the subject. Much to her surprise the reports did not have any detectable effect on citizens' knowledge. Later she fielded a third wave of the same survey. It followed a period of eight weeks in which the local news paper ran a weekly add on page one which highlighted city performance data. The data in the newspaper was often depicted with a drawing or cartoon. In the subsequent survey Roher could detect improvements on citizens' knowledge on a number of

performance dimensions. One possible interpretation of the finding is that performance information must be made more vivid and easily digestible for it to have an impact on citizens. Unfortunately, at the time her findings were largely ignored and governments continued to rely on dry government reports to communicate about performance with the public (Lee, 2006).

In summary, from the onset, those communicating performance information to the wider public have a very difficult task at hand: The very characteristics (i.e., factual and numerical) of the information they disseminate are not psychologically well suited for the recipient. It raises the natural question of how citizens then interpret performance numbers when they actually engage with them. In the following three themes of research which tackles this question are highlighted.

### *The Dominance of Negative Performance Information*

The first dimension is the valence of the information. In psychology there has been extensive research on the asymmetrical effect of negative and positive information. It is captured by the concept of the negativity bias which denotes an asymmetrical response to positive and negative information yielding the effect that “negative events are more salient, potent, dominant in combinations, and generally efficacious than positive events” (Rozin & Royzman, 2001, p. 297). This insight has provided important afterthought for some of the key assumptions underlying performance management. Here the implicit assumption has been that positive and negative information carry equal weight in the eye of citizens. Ideally then good and bad performance leads citizens to respond with equal amounts of praise and blame for organizations performing in either way. Prewitt (1987) captures the ideal: “When economic and social indicators are moving in politically popular directions, political credit is

claimed; when they are moving in unpopular directions, political blame is assigned. Here, then, is a contribution of public statistics to the workings of democracy.” At the same time the negativity bias underpins the theory of blame avoidance which often has been linked to managerial response to poor performance (Weaver, 1986, p. 373). If citizen weight negative performance more than positive information then it becomes a lot more important for policy makers and managers to avoid blame rather than get praise (Marvel, 2014, Moynihan, 2012, Hood 2007, Hood 2010, Dixon & Hood 2010; Charbonneau & Bellavance 2012; George et al., 2017; James et al. 2016; Olsen 2017c). The negativity bias has therefore become a central object of research in the response to performance information because it both changes our naïve assumptions about the symmetry of information effects and at the same time help explain some of the behavior and response that can be observed among managers.

Drawing on a broader political science literature on voting behavior (Bloom & Price, 1975, Kinder & Kiewiet, 1979, Mueller, 1973), a number of studies have looked at how incumbent politicians are punished for bad public service performance but are not rewarded for good performance (James & John, 2007, Boyne et al., 2009). These studies provide a clear skewness towards negative performance information in citizens’ responses. To get a closer look at the mechanism, studies have tried to disentangle the negativity bias from potential confounding effects. For instance, it may not be possible to observe negative and positive information of the same magnitude in the real world. Maybe negative information is more extreme and observed less frequently because public service often is moderately good or about average (Stipak, 1979). If that is the case then voters are not just responding to negative information but to a range of other relevant aspects of the information.

One solution to this has been to rely on equivalence framing, where “two logically equivalent (but not transparently equivalent) statements of a problem lead decision makers to choose different options” (Rabin, 1998, p. 36) are presented to citizens in an experiment. For

instance one might expose one group to the performance data showing that “10% of kids at a school *fail* their exam”, while others are being told that “90% of kids at a school *pass* their exam”. The two pieces of performance information are obviously logically equivalent: they refer to the same underlying level of performance. However, the former stresses negative aspects (“fail rate”) while the other one highlights positive aspects of performance (“pass rate”). One hereby makes sure that positive and negative information are not any different in extremity or some other aspect – only difference is the valence of the label. At a theoretical level, the valence shift in labels changes the encoding that the recipient makes of the information. A negative label leads to a negative encoding which increases the probability that negative associations are produced. For instance, if citizens are exposed to a certain “fail rate” at a school it becomes more likely that other negative aspects about school performance become salient. The opposite then happens when exposed to a positive frame. For instance, Olsen (2015a) exposes citizens to a patient satisfaction measure as well as a patient dissatisfaction. They were then asked to rate a hospital's performance based on the information. Their evaluation of a hospital with a 90% satisfaction was one standard deviation better than if told the hospital had a 10% dissatisfaction rate. This difference was comparable to very large changes in the actual performance of the hospital. Specifically, changing the valence of the labels was equivalent to changing the satisfaction rate with 22 percentage points.

By adding a neutral treatment that combines both negative and positive information, one can more carefully estimate the asymmetry between negative and positive performance information. Using these methods confirms a strong negativity bias in citizens' perception of public organizations based on performance data (Olsen 2015b). Citizens rate organizations at about the same level if exposed to neutral or positive performance information and only negative information drags down their ratings.

In sum, communication of performance to citizens will be largely contingent on the valence of the information. Negative information will have a larger impact and just changing the label while holding constant the level of performance will change perceptions of an organization dramatically. More generally, evaluations are highly asymmetrical in the sense that negative information leads to negative evaluations while neutral and positive information render very similar responses. Overall, it indicates that performance management will be affected by strong incentives to avoid blame while less attention should be diverted to get praise (Nielsen & Moynihan, 2017).

### *The Role of Performance Comparisons and Reference Points*

Performance information provides great detail and often carries implicit cues of precision and objectivity. As Stone (2012, pp. 191-196) noted this observation as well since “numbers, by seeming to be so precise, help bolster authority of those who count...numbers impart an aura of expertise and authority to the people who produce and use them.” At first hand, performance information swipes away the ambiguity and provides a sense of reliable information about an important topic: the performance of public organizations. However, a closer look reveals that even the most seemingly unambiguous and simple numerical performance information requires interpretation from the recipient. The previous section highlighted saw how the valence of the label attached to performance numbers can be a powerful cue when citizens interpret performance. Next attention is turned to how the communication of performance numbers is interpreted. That is how do citizens approach the task of making sense of performance numbers?

Psychology points to how individuals tend to focus on round numbers – or certain focal or symbolic values. For instance, powers and multipliers of 10 are found most frequently in

written material across time and cultures (Dehaene & Mehler, 1992, Dorogovtsev, Mendes, & Oliveira, 2006). These round numbers serve as “cognitive reference points” (Rosch, 1975). Studies also suggest that humans can be motivated to achieve results that satisfy symbolic values (Pope & Simonsohn, 2010). This indicates that individuals are likely to simplify numerical information whenever they rely on it for judgment and decision making. One way of approaching the numbers is by relying on the leftmost digit, that is, the first digit in the performance number. Studies of communication of prices show that consumers rely on a leftmost-digit-bias implying that consumers neglect or down weight digits that follow the left most digit (Thomas & Morwitz, 2005). For performance information, a similar effect has been identified for the case of school performance data. If citizens are offered simple two digit grade averages they will almost completely ignore the second digit and solely base their evaluation of a school on the first digit (Olsen, 2013). A similar finding has been found for citizens’ attitudes about budgetary decision based on numerical information about budget proposals (Malhotra & Margalit, 2010).

Often citizens are offered more than a single value of performance information that they will need to consider. Benchmarking and the use of comparative data has been a key issue since the early works of Herbert Simon who stressed that comparisons were needed for both citizens and managers to make valid inferences about the data. Comparisons have the upside of reshaping a question about absolute value (e.g., is X level of performance high or low?) into a relative question (e.g., is X level of performance better or worse than Y level for performance?). In short, comparisons allow citizens to substitute a difficult question with a far easier one.

For Simon (1939, p. 106) there was two main ways of doing performance comparisons: “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.”

This way of comparing matches later psychological research on how individuals can rely on either historical comparisons in time or social comparisons with others (Albert, 1977, Festinger, 1954). Social comparisons are comparisons to other similar units (e.g., comparisons with other municipalities, organizations, countries etc), while historical comparisons are comparisons over time for the same unit (e.g., comparisons with last year, previous quarters, or past decades etc.). In public management the research has focused on how citizens respond to relative vs. absolute performance comparisons and what type of comparisons has the most impact on impressions of a service (James 2011, Hansen, Olsen, & Bech, 2015, James & Moseley, 2014). Using actual performance reports, James (2011) social shows that comparisons of performance across units raise citizens' perceived performance and satisfaction with the services. If compared directly, studies find that both absolute and relative performance measures affect citizens' perceptions (James & Moseley, 2014). However, relative performance comparisons allow citizens to make more certain judgments without spending more time (Olsen, 2017, Mussweiler, 2003). Also, when directly compared, social comparisons have a larger effect on service perceptions than historical comparisons (Charbonneau & Van Ryzin 2015, Olsen, 2017).

### *Motivated Reasoning and Performance Information*

The final set of studies turn attention to something else than the performance information itself. It focuses instead on the interaction between the recipients' dispositions and beliefs and how they correspond to beliefs about the organization being evaluated (Kahan et al., 2013, Taber & Lodge, 2006, Petersen et al., 2012, Leeper & Slothuus, 2014). Even in the early studies of performance information we find the notion that citizens differ in the knowledge, interest and abilities to comprehend and rely on numerical data (Beyle, 1928, Roher, 1941a,

Roher, 1941b). However, with the theoretical work on motivated reasoning the idea is introduced that prior attitudes and predispositions can be a primary source of differences in information processing among citizens. The concept of motivated reasoning highlights that citizens' motivations, when confronted with performance data, can take many forms. Specifically, they can take the form of accuracy goals or directions goals (Kunda, 1990). Up until this point the assumption has been that citizens care about accuracy – they aim to arrive at a conclusion about the data which reflects the true state of performance. That is, focus has been on how various biases systematically skew these accuracy goals. However, with motivated reasoning the alternative hypothesis can also be that citizens are motivated to protect their prior beliefs, or what here is denoted “directional goals”. That is, citizens are motivated to interpret the data in a way that is consistent with the direction of their prior beliefs. The interest in this questions has only been strengthened by the fact that politicians have been found to strongly rely on prior attitudes when evaluating performance data and that the effect of these attitude are amplified even more when more data is provided (Bækgaard et al. 2017).

The most straight forward example is how citizens ideological beliefs about how private or public provision affect performance. Citizens' views on this topic have been found to vary widely and across many different performance dimensions (Andersen & Hvidman, 2016). The question then becomes how these predispositions filter the impression or performance evaluation derived from performance data that either go against or support prior beliefs about public or private provision.

Bækgaard and Serritzlew (2016) randomly assign citizens to performance information showing that either a public or private hospital is doing best in terms of avoiding maltreatment in operations. The performance data is communicated as frequencies and require that the recipient calculate maltreatment rates in order to answer the question of if the

private or public hospital is doing best (see also Kahan et al. (2017)). Interestingly, in a placebo condition the very same data is presented without any indication of whether the hospitals are private or public. The results show that citizens' prior beliefs about public and private provision affect their calculation of complication rates. However, the effect of prior beliefs vanishes when the data no longer indicates if the hospital is public and private. Along a similar line of argument, James and Van Ryzin (2016) look at how providing citizens with a prime to encourage party political thinking affects interpretations of data relative to a prime about their needs as service users. Using the case of Obamacare in the US, they show that under the political prime, Democrats rate evidence favorable to the healthcare program's performance more strongly than Republicans. If Democrats are given a choice of various performance information to inform their assessment of the program's performance, they tend to choose indicators that report more positive performance than those chosen by Republicans.

On a more general level, studies have highlighted that research needs to take citizens' general beliefs about public services into account when understanding how they process new information. For instance, using psychological implicit association test, Marvel (2015) has found that US citizens rely on unconscious and often negative attitudes towards public services. These can serve as a difficult filter for performance information without citizens making selective choices about what information to pay attention to and rely on. One example of this is provided by James and Van Ryzin (2015) who find that citizens who were exposed to information about good performance on a measure of customer satisfaction with the US Veterans Affairs and Citizenship and Immigration departments, evaluated the information as more credible when it was provided from the independent American Customer Satisfaction Index than when it came from the government agencies themselves. It clearly highlights the importance of understanding interactions between citizens' prior beliefs and

how they are supported or rejected by the content of the communicated performance data and the source of the data.

### **Conclusion and Future Research**

Performance management highlights the importance of communicating performance information to citizens. It hereby also highlights the need for performance management to engage in very direct dialogue with communication science. In recent years this interchange has unfolded with behavioral and experimental research on how performance information affects citizens' perception of public organizations and their attitudes towards different services. These studies help disentangle the different psychological cues that citizens draw when confronted with performance data.

However, it not only provides a more complete view of how citizens actually make sense of performance data; it also helps our understanding of how managers and politicians respond to performance measures. Examples include how the negativity bias in citizens' response to performance information increases the incentives of managers to use blame avoiding strategies. Ideally, the aim should be a research program in which micro-level studies of citizens are mirrored in meso-level studies of managers and organizations which in turn reflect macro studies on how performance information regimes change over time and vary between countries. Importantly, with a common psychological foundation for the study of citizens we can arrive at a unified global research agenda that draws on human cognition and emotions to understand the effects of performance data. As we then move on to the meso- and macro-level we can increasingly emphasize local context and variations.

This being said, the modern experimental and behavioral research on performance information also has a number of blind spots and leaves certain questions unanswered:

First, these studies provide a clear micro-level perspective by studying the cognitive processes that shape how citizens make sense of performance data. This view has the advantage of providing clear information about the implications for how certain ways of communicating performance will affect citizens' perceptions. It hereby provides correspondence between information content and effects among recipients. However, the question is how to integrate these insights with macro-level studies of how performance management. Much of the existing research stresses differences in performance management and measurement practices between countries and sectors. The experimental and behavioral approach to performance information stresses similarities across contexts stemming from the fundamental cognitive processes that are studied. The key question moving forward should therefore be: How can micro- and macro-level studies of performance measurement be integrated?

Second, the behavioral and experimental agenda in performance information research is to date a collection of loosely coupled facts about how citizens make sense of performance information. These are all very appealing individually as they provide clear predictions effects and guidance to practitioners that want to understand how performance information works. However, apart from a common grounding in social and cognitive psychology, the different core findings do not refer to a unified or coherent theoretical model. It raises some questions about how these various biases fit together and how they might boost or cancel out each other. Model building aimed at integrating these different insights should therefore have a high priority. Here may neighboring fields like communication studies and political psychology become handy, as they both have taken aim at building more elaborate models of human information processing which combines insights from a diverse set of theories. The motivating question moving forward should therefore be: How can findings in behavioral performance information studies be refined in a comprehensive theoretical model?

Third, a potential way forward for both theoretical and methodological integration is to view the reception of performance information as an interdisciplinary study where scholars outside of public management have a relevant contribution to make. Today performance information is mainly studied by scholar with a background in public administration and political science. However, in other disciplines highly relevant studies can be found that clearly speaks to the topic of public sector communication of performance. Examples include studies of hospital performance data in medicine (Hibbard et al., 2003, Hibbard et al., 2005), studies of campaigns and media in political communication and political psychology (Gross, 2008), and the study of economic voting or performance voting in political behavior (Lau 1982, Lau 1984). There is at task at hand for future research to cut across disciplinary boundaries and embrace outside insights. Here lies a leapfrog potential that allows scholars in public administration to benefit from strong evidence derived in other fields. The motivating question moving forward should therefore be: What existing evidence does other fields already possess that is not currently recognized as a potential contribution to the study of performance information?

Fourth, most of the studies reviewed here rely on survey or lab experiments in which respondents are exposed to real performance data or hypothetical vignettes. These methods impose restrictions on the external and ecological validity of our knowledge of performance information. For instance, how do the different effects play out in settings where citizens are exposed to multiple and potentially ambiguous sources of information about government performance? The natural response is to conduct more field experiments in which citizens are exposed to government performance information in their natural environment. The motivating question moving forward should therefore be: What is the external and ecological validity of the recent behavioral and experimental research on citizens' inferences from performance data?

Fifth, the final unanswered question is about the ethics of studying the communication of performance information to citizens. A methodological move to field experiments also increases the importance of clear ethical guidance. Many of the insights obtained from these studies can potentially be used to manipulate and skew the presentation of performance data in order to highlight certain aspects of performance. For instance, managers can choose to present performance with comparative benchmarks that make their own performance look more favorable, or they might rely on insights from equivalence framing and use labels that shed their organization in a positive light. It raises a question about the ethics and misuse of experimental and behavioral research on communication performance information. Many ways of presenting performance information can seem very subtle and will often pass as technical details, but they can indeed have a powerful impact. This calls for more observation studies into manipulation and gaming of performance data. It also highlights a need for update bureaucratic values and norms to take into account that a lot of communication with citizens today takes the form of numerical performance data. The motivating question moving forward should therefore be: What are the ethical concerns of studying the details of how performance information affect citizens and how professional norms be developed for the communication of performance data?

## References

- Andersen, S. C., & Hjortskov, M. (2016). Cognitive Biases in Performance Evaluations. *Journal of Public Administration Research and Theory*, 26(4), 647–662.
- Askim, J. (2007). How do politicians use performance information? An analysis of the Norwegian local government experience. *International Review of Administrative Sciences*, 73(3), 453–472.
- Beyle, H. C. (1928). Objective Studies as a Basis for Functionalized Governmental Reporting. *Proceedings of the Annual Meeting of the Governmental Research Association*. New York: Governmental Research Association.
- Beyle, H. C. (1936). Checking Response to Municipal Publicity. *Public Management*, 18(6), 163-66.

- Blom-Hansen, J., Morton, R., & Serritzlew, S. (2015). Experiments in public management research. *International Public Management Journal*, 18(2), 151–170.
- Boyne, G. A., James O., John, P., & Petrovsky, N. (2009). Democracy and Government Performance: Holding Incumbents Accountable in English Local Governments. *Journal of Politics* 71(4), 1273–1284.
- Bækgaard, Marti, & Serritzlew, S. (2016). Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension. *Public Administration Review*, 76(1), 73–82.
- Bækgaard, M., Christensen, J., Dahlmann, C. M., Mathiasen, A., & Petersen, N. B. G. (2017). The role of evidence in politics: Motivated reasoning and persuasion among politicians. *British Journal of Political Science*, 1–24.
- 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.
- Dehaene, S., & Mehler J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, 43, 1–29.
- Dorogovtsev, S. N., Mendes, J. F. F., & Oliveira, J. G. (2006). Frequency of occurrence of numbers in the World Wide Web. *Physica A: Statistical Mechanics and Its Applications*, 360(2), 548–556.
- Espeland, W. N., & Sauder, M. (2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology*, 133 (1), 1–40.
- George, B., Desmidt, S., Nielsen, P. A., & Bækgaard, M. (2017). Rational planning and politicians' preferences for spending and reform: replication and extension of a survey experiment. *Public Management Review*, 19(9), 1251-1271.
- Grimmelikhuijsen, S., Jilke, S, Olsen, A. L., & Tummers, L. (2017). Behavioral Public Administration: Combining Insights from Public Administration and Psychology. *Public Administration Review*, 77(1), 45–56.
- Gross, K. (2008). Framing persuasive appeals: Episodic and thematic framing, emotional response, and policy opinion. *Political Psychology*, 29, 169–192.
- Hansen, Kasper M., Olsen, A. L., & Bech, M. (2015). Cross-National Yardstick Comparisons: A Choice Experiment on a Forgotten Voter Heuristic. *Political Behavior*, 37(4), 767–789.
- Hatry, H.P. (1999). *Performance Measurement: Getting Results*. Washington DC: Urban Institute Press.
- Herbst, S. (1993). *Numbered voices: How opinion polling has shaped American politics*. University of Chicago Press.
- Hibbard, J. H., Stockard, J., & Tusler, M. (2003). Does Publicizing Hospital Performance Stimulate Quality Improvement Efforts? *Health Affairs*, 22(2), 84–94.
- Hibbard, J. H., Stockard, J., & Tusler, M. (2005). Hospital Performance Reports: Impact on Quality, Market Share, and Reputation. *Health Affairs*, 24(4), 1150–60.

- Hvidman, U., & Andersen, S. C. (2016). Perceptions of Public and Private Performance: Evidence from a Survey Experiment. *Public Administration Review*, 76(1), 111-120.
- Hood, C. (2007). What happens when transparency meets blame-avoidance? *Public Management Review*, 9(2), 191-210.
- Hood, C. (2010). *The Blame Game: Spin, Bureaucracy and Self-Preservation in Government* Princeton: Princeton University Press.
- Hood, C., & Dixon, R. (2015). *A government that worked better and cost less? Evaluating three decades of reform and change in UK central Government*. Oxford: Oxford University Press.
- Holzer, M., & Yang, K. (2004). Performance Measurement and Improvement: An Assessment of the State of the Art. *International Review of Administrative Sciences*, 70(1), 15–31.
- Hvidman, U., & Andersen, S. C. (2016). Perceptions of Public and Private Performance: Evidence from a Survey Experiment. *Public Administration Review*, 76(1), 111–120.
- James, O., & John. P. (2007). Public Management at the Ballot Box: Performance Information and Electoral Support for Incumbent English Local Governments. *Journal of Public Administration Research and Theory*, 17(4), 567–80.
- James, O. (2011a). Performance Measures and Democracy: Information Effects on Citizens in Field and Laboratory Experiments. *Journal of Public Administration Research and Theory*, 21, 399–418.
- James, O. (2011b). Managing Citizens' Expectations of Public Service Performance: Evidence from Observation and Experimentation in Local Government. *Public Administration*, 89(4), 1419–35.
- James, O. & Moseley, A. (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.
- James, O. & Van Ryzin, G.G. (2015). Incredibly good performance: An experimental study of source and level effects on the credibility of government. *American Review of Public Administration*, 47(1), 23–35.
- James, O., & Van Ryzin, G. (2016). Motivated Reasoning about Public Performance: An Experimental Study of How Citizens Judge Obamacare. *Journal of Public Administration Research and Theory*, 27(1), 197–209.
- James, O., Gilke, S., Petersen, C. & Van de Walle, S. (2016). Citizens' Blame of Politicians for Public Service Failure: Experimental Evidence about Blame Reduction through Delegation and Contracting. *Public Administration Review*, 76(1), 83–93.
- James, O. & Olsen, A. L. (2017). Citizens and Performance Measures: Making Sense of Performance Information. In: James, O., Gilke, S., & Van Ryzin G. G. (Eds.). *Experiments in Public Management Research: Challenges and Contributions* (pp. 270-290). Cambridge, UK: Cambridge University Press.

- Johnsen, Å. (2012). Why Does Poor Performance Get So Much Attention in Public Policy? *Financial Accountability & Management* 28(2), 121–142.
- Kahan, Dan M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated Numeracy and Enlightened Self-Government. *Behavioural Public Policy*, 1, 54–86.
- Kroll, A. (2013). The other type of performance information: Non-routine feedback, its relevance and use. *Public Administration Review*, 73(2), 265–76.
- Kunda, Z. (1990). The Case for Motivating Reasoning. *Psychological Bulletin*, 108(3), 480–98.
- Lacetera, N., Pope, D. G., & Sydnor, J. R. (2012). Heuristic Thinking and Limited Attention in the Car Market. *American Economic Review*, 102(5), 2206–2236.
- Lau, R. R. (1985). Two Explanations for Negativity Effects in Political Behavior. *American Journal of Political Science*, 29(1), 119–138.
- Lau, R. R. (1982). Negativity in Political Perception. *Political Behavior*, 4(4), 353–377.
- Lavertu, S. (2016). We all need help: “Big data” and the mismeasure of public administration. *Public administration review*, 76(6), 864-872.
- Lee, M. (2006). Empirical experiments in public reporting: Reconstructing the results of survey research, 1941–42. *Public Administration Review*, 66(2), 252–262.
- Leeper, T. J., & Slothuus, R. (2014). Political Parties, Motivated Reasoning, and Public Opinion Formation. *Political Psychology*, 35 (S1), 129–156.
- Levin, I. P., Schneider, S. L., & Gaeth, Gary J. (1998). All Frames are not Created Equal: A Typology and Critical Analysis of Framing Effects. *Organizational behavior and human decision processes*, 76(2), 149–188.
- Levin, I. P., & Gaeth, G. J. (1988). How Consumers are Affected by the Framing of Attribute Information Before and After Consuming the Product. *Journal of Consumer Research*, 16, 374–378.
- Lodge, M., & Taber, C.S. (2013). *The Rationalizing Voter*. Cambridge: Cambridge University Press.
- Margetts, H. Z. (2011). Experiments for public management research. *Public Management Review*, 13(2), 189–208.
- Malhotra, N., & Margalit, Y. (2010). Short-term communication effects or longstanding dispositions? The public’s response to the financial crisis of 2008. *Journal of Politics*, 72(3), 852–867.
- Marvel, J. D. (2015). Unconscious bias in citizens’ evaluations of public sector performance. *Journal of Public Administration Research and Theory*, 26 (1), 143–158.
- Marvel, J. D. (2014). The Boston Marathon Bombings: Who's to Blame and Why It Matters for Public Administration. *Public Administration Review*, 74(6), 713–725.

- Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. *Public Administration Review*, 76(6), 928-937.
- Moynihan, D. P. (2008). *The dynamics of performance management: Constructing information and reform*. Washington, DC: Georgetown 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*, 20(4), 849–866.
- Moynihan, D. P. (2012). Extra-Network organizational reputation and blame avoidance in networks: The Hurricane Katrina example. *Governance*, 25(4), 567–588.
- Moynihan, D. P. (2016). Political use of performance data. *Public Money & Management*, 36(7), 479–481.
- Nielsen, P. A., & Moynihan, D. P. (2017). How Do Politicians Attribute Bureaucratic Responsibility for Performance? Negativity Bias and Interest Group Advocacy. *Journal of Public Administration Research and Theory*, 27(2), 269–283.
- Olsen, A. L. (2013a). Leftmost-digit-bias in an enumerated public sector? An experiment on citizens' judgment of performance information. *Judgment and Decision Making*, 8(3), 365–371.
- Olsen, A. L. (2013b). The politics of digits: evidence of odd taxation. *Public Choice* 154 (1-2), 59–73.
- Olsen, A. L. (2015a). Citizen (Dis)Satisfaction: An Equivalence Framing Study. *Public Administration Review* 75(3), 469–478.
- Olsen, A. L. (2015b). Simon Said, We Didn't Jump. *Public Administration Review*. 75(2), 469–478.
- Olsen, A. L. (2015c). The Numerical Psychology of Performance Information – Implications for Citizens, Managers, and Policy Makers. *Public Performance & Management Review*, 39(1), 100–115.
- Olsen, A. L. (2015d). Negative Performance Information Causes Asymmetrical Evaluations and Elicits Strong Responsibility Attributions. *111th Annual Meeting of the American Political Science Association San Francisco, September, 2015*.
- Olsen, A. L. (2017a). Human Interest or Hard Numbers? Experiments on Citizens' Selection, Exposure, and Recall of Performance Information. *Public Administration Review*, 77(3), 408–420.
- Olsen, A. L. (2017b). Compared to What? Experimental Evidence on Social and Historical Reference Points in Performance Evaluation. *Journal of Public Administration Research and Theory*, 27(4), 562–580.
- Olsen, A. L. (2017c). Responding to Problems: Actions are Rewarded Independent of the Outcome, *Public Management Review*, 19(9), 1352–1364.
- Osborne, D., & Gaebler, T. (1992). *Reinventing government: How the entrepreneurial spirit is transforming government*. Reading MA: Adison Wesley Public Comp.

- Perrin, B. (1998). Effective use and misuse of performance measurement. *American Journal of Evaluation*, 19(3), 367–79.
- Petersen, M. B., Skov, M., Serritzlew, S., & Ramsøy, T. (2012). Motivated Reasoning and Political Parties: Evidence for Increased Processing in the Face of Party Cues. *Political Behavior*, 35 (4), 831–854.
- Pope, D., & Simonsohn, U. (2011). Round numbers as goals evidence from baseball, SAT takers, and the lab. *Psychological science*, 22(1), 71–79.
- Rabin, M. (1998). Psychology and Economics. *Journal of economic literature*, 36(1), 11–46.
- Radin, B. (2006). *Challenging the Performance Movement: Accountability, Complexity, and Democratic Values*. Washington, DC: Georgetown University Press.
- Roher, M. (1941a). Education of a Citizen. *National Municipal Review*, 30(4), 192–198.
- Roher, M. (1941b). How Good Are Reports to Citizens. *Western City*, 17(8), 17–9.
- Rozin, P., & Royzman, E. B. (2001). Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review* 5(4), 296–320.
- Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, 7, 532–547.
- Slovic, P. (2007). If I Look at the Mass I Will Never Act: Psychic Numbing and Genocide. *Judgment and Decision Making*, 2(2), 79–95.
- Soroka, S N. (2006). Good News and Bad News: Asymmetric Responses to Economic Information. *Journal of Politics* 68(2), 372–385.
- Stone, D. A. (1997). *Policy paradox: The art of political decision making*. New York: WW Norton.
- Simon, H. A. (1937). Comparative statistics and the measurement of efficiency. *National Municipal Review*, 26(11), 524–27.
- 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. *Quarterly Journal of Economics*, 69(1), 99–118.
- Taber, C. S., & Lodge, M. (2006). Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*, 50(3), 755–69.
- Thomas, P. (2006). *Performance Measurement, Reporting, Obstacles and Accountability: Recent Trends and Future Directions, Australia and New Zealand School of Government Monograph*. Canberra: ANU E-Press
- Thomas, M., & Morwitz, V. (2005). Penny wise and pound foolish: The left-digit effect in price cognition. *Journal of Consumer Research: An Interdisciplinary Quarterly*, 32(1), 54–64.
- Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4), 1039–1061.

Tummers, L., Olsen, A. L., Jilke, S., and Grimmelikhuijsen, S. (2016). Introduction to the Virtual Issue on Behavioral Public Administration. *Journal of Public Administration Research and Theory*, Virtual Issue (3), 1–3.

Van Dooren, W., Bouckaert, G., & Halligan J. (2010). *Performance management in the public sector*. New York & London, UK: Routledge.

Van Ryzin, G. G., & Lavena, C. F. (2013). The credibility of government performance reporting. *Public Performance & Management Review*, 37(1), 87–103.

Weaver, R. K. (1986). The Politics of Blame Avoidance. *Journal of Public Policy*, 6(4), 371–398.

Zillmann, D. (2006). Exemplification effects in the promotion of safety and health. *Journal of Communication*, 56(1), 221–237.