International Public Management Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/upmj20

Factoring in the human factor: experimental evidence on how public managers make sense of performance information

Nicola Belle, Giorgio Giacomelli, Sabina Nuti & Milena Vainieri

To cite this article: Nicola Belle, Giorgio Giacomelli, Sabina Nuti & Milena Vainieri (2022) Factoring in the human factor: experimental evidence on how public managers make sense of performance information, International Public Management Journal, 25:2, 178-191, DOI: 10.1080/10967494.2021.1975856

To link to this article: https://doi.org/10.1080/10967494.2021.1975856

INTERNATIONAL PUBLIC MANAGEMENT JOURNAL
2022, VOL. 25, NO. 2, 178–191
https://doi.org/10.1080/10967494.2021.1975856

Factoring in the human factor: experimental evidence on how public managers make sense of performance information

Nicola Belle[a], Giorgio Giacomelli[a,b], Sabina Nuti[a], and Milena Vainieri[a]

[a]Management and Healthcare Laboratory, Institute of Management and Department EMbeDS, Scuola Superiore Sant'Anna, Pisa, Italy; [b]SDA Bocconi School of Management, Milano, Italy

ABSTRACT

We conducted a randomized controlled trial with 902 public hospital department heads to investigate whether and how their perceived understanding and objective recall of performance information may depend on the way it is presented, specifically, the features of its graphical representation. Participants were randomly assigned to one of three alternative graphical displays of the same data, namely a target, a radar, or a pathway chart. We then measured both subjects' perceived understanding—i.e., how clear and understandable they found the performance metrics—and their actual information recall—as assessed by an objective test. Our analyses demonstrate that alternative visual representations of equivalent data may cause significant differences in public managers' ability to make sense of performance information. In the context of our experimental test, the pathway chart outperformed the other two in both perceived understanding and objective information recall, which suggests that public managers may have a good sense of which graphic information formats work best for them. We discuss how our findings can contribute to the advancement of behavioral public performance research.

ARTICLE HISTORY
Received 23 July 2020
Accepted 30 August 2021

Introduction

Measuring performance has been at the top of the agenda for public management for several decades (Arnaboldi, Lapsley, and Steccolini 2015; Johnsen 2005) (Moynihan 2008) and is a particularly longstanding tradition in the healthcare sector (Smith 2005). Over time, the ever-increasing complexity of public services has been mirrored by a tendency to create performance measurement systems (PMSs) that have become more and more technically complex, thus leading to dynamics that have been referred to as the "logic of escalation" (Pollitt 2013; Woelert 2015). With the use of this logic, performance management may become increasingly control-focused and consequential (Pollitt 2013), based on a deterministic reliance on PMSs. In addition, traditional control measurement systems have been complemented by multidimensional performance measures, in order to ensure comprehensive information that conveys both financial and non-financial information (Modell 2012; Woods and Grubnic 2008). In order to allow organizational decision-makers to cope with this level of information complexity, performance representation recently has tried to ease information selection by graphically displaying performance targets/trends and systematic benchmarks and adopting user-friendly reporting systems.

CONTACT Nicola Belle nicola.belle@santannapisa.it Management and Healthcare Laboratory, Institute of Management and Department EMbeDS, Scuola Superiore Sant'Anna, Pisa, Italy.

Supplemental data for this article is available online at https://doi.org/10.1080/10967494.2021.1975856.

Considering that one of the primary roles of PMSs is to support strategy within an organization (Franco-Santos et al. 2007), it appears clear that the way in which data are represented constitutes a feature of the PMS design that is highly consequential for the use of performance metrics. The need to spread out a high number of indicators that provide a comprehensive picture of and align with the strategies of the organization challenges users' limited cognitive capacity to acquire and process information. Keeping in mind that decision-making should be viewed as a boundedly rational process (Simon 1955), the design of a PMS—in particular how performance is displayed—may benefit from psychological insights (Grimmelikhuijsen et al. 2017). The emergence of "systematic biases that separate the beliefs that people have and the choices they make from the optimal beliefs and choices assumed in rational-agent models" (Kahneman 2003) urges choice architects (Thaler, Sunstein, and Balz 2013) to design decision-making tools to facilitate a purposeful use of information (Kroll 2015). In this view, the design of PMSs is in all respects a choice architecture (Thaler, Sunstein, and Balz 2013), as it implies the need to define information features that can affect public managers' comprehension, decisions, and behaviors.

With this perspective, James et al. (2020) recently suggested a new approach—public behavioral performance—that aims to connect two revolutionary changes that occurred in parallel and that have profoundly transformed public administration practice and research. On the one hand, over the last decades, the performance movement has fostered the widespread adoption of performance metrics by public organizations at all levels of government. More recently, a second revolution has occurred in the flourishing of research that combines insights from psychology with the use of experimental methodologies to investigate how public managers and employees (e.g. Cantarelli, Belle, and Belardinelli 2020; Fuenzalida, Ryzin, and Olsen 2021), politicians (e.g. Baekgaard et al. 2019; Desmidt and Meyfroodt 2021), and citizens (e.g. Olsen 2015a, 2017a, 2017b; Ryzin 2013) make sense of performance information. In connecting these two advances, public behavioral performance suggests "lessons to improve design and use of performance metrics in public management and democratic accountability" (James et al. 2020, 1) by taking into account supposedly irrelevant factors that may indeed shape individuals' perceptions, judgments, decisions, and behaviors. These factors include, for instance, the characteristics of numbers (Brunell and Glazer 2001; Olsen 2013, 2015b), frames (Belardinelli et al. 2018), and reference points (Belardinelli, Bellé, and Cantarelli 2020; Bellé, Cantarelli, and Belardinelli 2018; Olsen 2017b). Our study aims to contribute to the literature on public behavioral performance by experimentally investigating how public managers' perceived understanding and objective recall of performance information may depend on the way it is presented to them, specifically, the features of its graphical representation. Although the visualization of public performance data is a fundamental form of framing, this factor has so far been largely overlooked. One of the few exceptions is the work of Allwood, Hildon, and Black (2013) on the visualization of data; however, the authors focused mainly on specific performance indicators. More complex visualization tools have only been described (Matheus, Janssen, and Maheshwari 2020; Nuti, Seghieri, and Vainieri 2013), but not empirically tested for effectiveness. The few studies that analyzed the strengths and weaknesses of alternative graphical representations of performance information (Nuti et al. 2018; Stafoggia et al. 2011), mainly have adopted observational methodologies. This article provides some of the first experimental investigations of the causal effect that alternative data visualizations can have on public professionals' understanding and recall of performance information. As such, it offers empirical evidence on an important but understudied topic. Specifically, our findings can help pave the way for more focused and informed use of public performance data by showing which types of visualizations are most effective in helping public professionals process data accurately.

Performance information dynamics in the context of healthcare

Healthcare organizations are complex and adaptive systems characterized by various organizational designs, multiple vertical and horizontal interconnections, and a high level of formal

control. Added to this is a high degree of individual professional autonomy and influence, with a great part of the expenditure governed by clinicians using knowledge-driven, and yet discretionary, choices (e.g., drug prescriptions or surgical procedures). This implies a peculiar reverse power structure, where a traditional hierarchy is unlikely to be effective (Dickinson et al. 2016). Overall performance in this environment results from the interdependent work of highly specialized professionals and managers, who all contribute to creating value across the organization. In this context, communicating information is a key mechanism to foster the functioning of the "reverse hierarchy" model (Vainieri et al. 2019). However, as no communication is neutral (Thaler, Sunstein, and Balz 2013), the ways in which information is presented to decision-makers is likely to affect their choices and alter their behaviors.

This study explored how the way in which performance information is presented to physicians in senior management positions within public hospitals may affect their ability to make sense of that information. We conducted a randomized controlled trial (RCT) with a sample of chief physicians who were randomly assigned to one of three different graphical representations of the same performance indicators: a radar chart, a "dartboard" (Nuti et al. 2012), or a pathway-like chart (Nuti et al. 2018). After having been exposed to one of the three performance information formats for the same amount of time, subjects answered the same two batteries of questions. The first set of questions elicited participants' perceptions about information clarity and subjective understanding (i.e., how much they thought they had understood). The second battery of questions tested objective recall and understanding of the performance information.

Our study contributes to both theory and practice. From a theoretical standpoint, we add novel experimental evidence to a nascent stream of research that explores the effects of framing on how public managers make sense of performance information (Bellé, Cantarelli, and Belardinelli 2018). We also address the relevance for practice by illuminating the implications of alternative graphical representations of equivalent performance indicators for the management of a highly educated public workforce.

The manuscript is organized as follows. Section 2 reviews the literature on data presentation in healthcare and introduces some cornerstones of behavioral science for public sector studies; the hypotheses that will be tested throughout the empirical investigation are also formulated. Section 3 describes the study methodology. The results of our statistical analyses are then reported in Section 4, which also reviews the main limitations of our research. In Section 5, we discuss the results of our randomized controlled trial and provide concluding remarks.

Theory

Data visualization and presentation of performance information

Scholars have long been discussing the aims of PMSs and have thoroughly debated differences and interconnections between performance measurement and management (Radnor and Barnes 2007; Van Dooren, Bouckaert, and Halligan 2015). However, the comparison of alternative performance information formats has been largely disregarded in public sector research until recent times (Fryer, Antony, and Ogden 2009). Indeed, data displays traditionally have been chosen on the basis of common sense rather than evidence (Hildon, Allwood, and Black 2012a). This is problematic because there is no such thing as a neutral presentation of performance data and the choice among alternative visualizations may fundamentally alter the information-decision continuum (Rahman, Adamu, and Harun 2017). Research in this area recently has gained traction, particularly due to the spread of big data analytics (Grover et al. 2018). This trend is also influencing public sector research, with an increasing number of studies across different areas and disciplines (Isett and Hicks 2018; Matheus, Janssen, and Maheshwari 2020; Stadler et al. 2016).

Three main research streams emerge from Lindquist's (Lindquist 2011b) overview of extant scholarship on data visualization: information visualization and data analytics, graphics and information display; and visual facilitation for thinking and strategy. The first approach "emerged out of computing and graph-making, motivated by the need to visually represent increasingly large data-sets" and "to enhance how humans can analyze and learn from this information" (Lindquist 2011b, p. 2). The second area of research deals with the design and representation of data for the purposes of communication and marketing. Lastly, the third area of study aims primarily to help decision-makers to better understand and grapple with complex challenges, aided by visual solutions that reduce cognitive load (Huang, Eades, and Hong 2009), offload short-term memory, and facilitate inferences (Tufte 1990). Despite little evidence for how visualization shapes policy decisions (Isett and Hicks 2018; Lindquist 2011a), general support for the enhanced efficacy of data visualization, compared to other forms of information presentation, has been highlighted by the increasing number of studies (Bresciani and Eppler 2009; Olsen 2017a; Slingsby et al. 2014).

In health care management scholarship, several authors recently have addressed the topic of understanding performance information by comparing the efficacy of different visual displays for comprehension by either patients (Hildon, Allwood, and Black 2012b) or professionals (Allwood, Hildon, and Black 2013; Stadler et al. 2016; Stafoggia et al. 2011). When it comes to the impact of presentational choices on professionals' information understanding, the literature agrees on the need for a presentation of data that is concise and minimizes cognitive effort (Dowding et al. 2015) but have produced inconsistent results for the relative efficacy of different displays. For instance, there is no unanimous consent on the possible advantages and drawbacks of information understanding when providing data in benchmarking (Morales-Silva et al. 2020; Zwijnenberg et al. 2016), just as it is ambiguous what matters relatively more between the visual display and the individual characteristics of professionals, such as preexisting numeracy and graph literacy (Dowding et al. 2015).

In order to unravel the ambiguities about the causal relationships between data visual display and information understanding, some of these studies draw on framing effects (Tversky and Kahneman 1981) as a theoretical lens through which to investigate the impact of presentational choices on information understanding and decision-making processes (Edwards et al. 2001; McNeil et al. 1982).

Equivalence framing and performance information

The bounded-rationality literature has long shown that individual judgment is systematically affected by cognitive limitations and by information availability and tractability (Simon 1955). Later work (Tversky and Kahneman 1974) has explored the psychology of intuitive beliefs and choices, that is "thoughts and preferences that come to mind quickly and without much reflection" (Kahneman 2002). Within this stream of studies, scholars have devoted great attention to framing effects, which have been defined by Kahneman as "discrepancies between choice problems that decision makers, upon reflection, consider effectively identical" (Kahneman 2002, p. 457). Such effects are relevant for performance information because the same piece of information may be presented in different ways without changing its content. An equivalence framing effect occurs when individuals respond in systematically different ways to objectively equivalent pieces of information that are framed differently (Levin, Schneider, and Gaeth 1998; Tversky and Kahneman 1981). From this follows:

Hypothesis 1. Different graphical representations of equivalent performance information will affect perceived clarity and understanding of that information.

Hypothesis 2. Different graphical representations of equivalent performance information will affect objective recall and understanding of that information.

Method

Study setting, design and participants

Participants in our experiment were 1,502 professionals with managerial responsibilities, all working at Health Authorities (HAs) of a large Regional Health System (RHS) in Northern Italy. Italian HAs are responsible for planning and delivering health care and social services to citizens in a specific geographic area. The Head of the Health Projects Unit of the RHS validated the final survey before agreeing to begin the data gathering step. The sample was composed of 1,502 respondents, with a 56% average response rate: respondents were invited to take part in the experimental survey on a voluntary basis, after completing a questionnaire about organizational climate, which itself collected 2,696 respondents (of a total population of 11,035). Of the 1,502 sample, 902 professionals were further selected to rule out previous knowledge of one of the visual formats. As the RHS has participated in an Inter-Regional Performance Evaluation System (IRPES) since 2014 (Nuti et al. 2016), it was possible that the respondents had some prior acquaintance with one of the performance visual formats, i.e. the dartboard. In order to neutralize this possible bias, respondents were asked at the beginning of the experimental survey to declare their prior acquaintance with the dartboard system, picking one among 'none', 'poor', 'intermediate' and 'good'; only the 902 professionals declaring no previous knowledge of the dartboard system were considered in the final sample. We administered our experimental survey through Qualtrics.

Understanding of performance information via different presentational formats

Our experiment consisted of displaying the same nine key performance indicators (KPIs) for a specific health pathway (namely, the maternity pathway) using three different graphical formats: a *target chart* (a.k.a."*dartboard*"), a *radar chart*, and a *pathway chart* (Nuti et al. 2018). The choice of those three experimental formats was made in close collaboration with and validated by the panel of experts responsible for the development and implementation of the performance measurement system shared by the Inter-Regional Performance Evaluation System (IRPES). To maximize realism, we selected the three graphical representations that the experts responsible for IRPES indicated as the three primary formats considered to become the standard within the regional health systems participating in the network.

Each of the three charts, unlike a simple pie or bar chart, represents a complex data visualization. All three charts convey a lot of information in a single picture. In particular, while the radar chart is a very common graphical representation in all sectors, the target chart was originally introduced in the first decade of 2000 in some Italian Regions of the healthcare sector in Nuti, Seghieri, and Vainieri (2013) and then recently has been used in more and more countries and official documents (see for instance, the country profile of the Health at Glance of the OECD, the EU State of Health reports, the health quality and safety commission in New Zealand) to represent overall performance of health care organizations. Finally, the third chart, the pathway chart uses the metaphor of the musical stave. It has been recently introduced in the Italian Regions (Nuti et al. 2018) with the specific aim to provide an overall picture of the performance of the healthcare pathways through notes (indicators) and musical bars (the phase of the pathway).

All three charts displayed the same set of nine indicators (the KPIs) selected from the maternal care performance assessment (Murante, Nuti, and Matarrese 2015):

- Late access to antenatal care for foreign women
- Participation by resident nulliparous in antenatal classes
- Recommendations of the family care center
- NTSV (Nulliparous, Term, Singleton, Vertex) cesarean sections

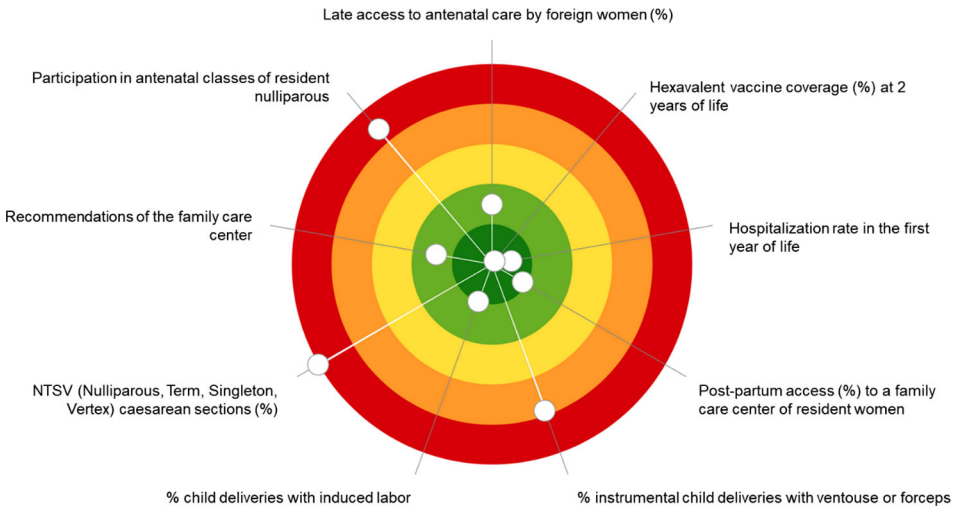

Figure 1. Target chart ("Dartboard").

- Deliveries with induced labor
- Instrumental deliveries with vacuum or forceps
- Post-partum access to a family care center of resident women
- Hospitalization in the first year of life
- Hexavalent vaccine coverage at 2 years of life

KPIs of the maternity pathway were chosen because they are usually more understandable and generally familiar than many of those concerning other healthcare pathways, regardless of the professional's profile and specialty.

The three alternative graphical techniques were applied as follows:

- *Target chart ("Dartboard")*. Each of the nine KPIs was depicted as a dot in a target-like display. The center of the target represented optimal performance, and the periphery represented poor performance. The display was divided in five color bands, with the highest scores (dark-green band) positioned in the center and the lowest ones (red band) in the outer circle (Figure 1).
- *Radar chart*. Each of the nine KPIs was represented on a different radius of a radar plot, where the length of each radius was proportional to the value of the represented KPI. The values were connected to form an enclosed area, with the points closest to the periphery representing good performance and the points in proximity to the center standing for poor performance. This presentational format is the actual reverse of the logic underlying the dartboard, with the exception of the use of colors (Figure 2).
- *Pathway chart*. The nine KPIs were displayed on a horizontal chart divided in five color bands, from red to dark-green. The KPIs were positioned on the chart according to the different sequential phases that the patients cross along the maternity pathway, thus relying on the metaphor of the musical stave, i.e. the set of horizontal lines and spaces used in sheet music (Nuti et al. 2018). This is intended to hint a message of continuity and patient-centeredness of the health pathway, and to allow users to focus on the strengths and weaknesses that characterize the service delivery in the different phases (Figure 3).

The three presentational formats were each randomly assigned to one of three groups of participants and displayed for a fixed amount of time (i.e., 90 seconds). After that, participants were

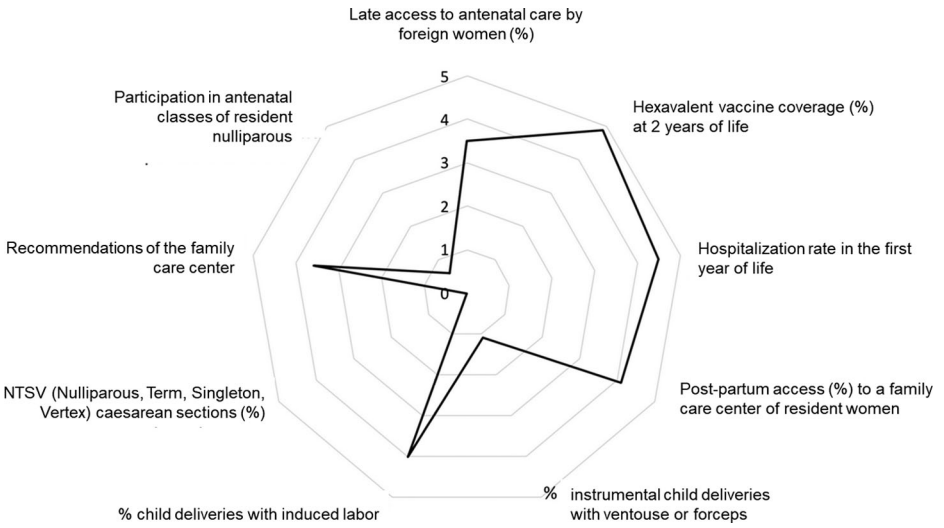

Figure 2. Radar chart.

asked to respond to two sets of questions concerning perceived understanding and objective recall of the data displayed. The two sets of questions were shown to the subjects on separate pages after the page with the graph, with no possibility to go back. The first set of questions regarded self-declared perceptions about *clarity of the graph*, *ease of reading*, *ease of recall*, *readability of the KPIs labels* and *clarity of phases*. These items were measured with a 5-point Likert-type scale, and respondents were asked to position themselves between total disagreement (1) and total agreement (5) with the content of the items. The second set of questions tested understanding of the performance displayed: respondents were asked to pick an answer out of 4 options about the *best performing phase*, the *worst performing KPI of a certain phase* and the *best performing KPI of a certain phase*.

Statistical analyses

In order to assess the validity of our hypotheses, five ordinary least squares (OLS) regression models were tested and the main effects of the experimental treatments were explored for each sub-dimension of perceived understanding (*clarity of the graph*, *readability of the graph*, *ease of recall*, *readability of KPI labels*, *and clarity of the phases*). Regarding *objective understanding*, on the other hand, we tested three logistic regression models to assess the main effects of the experimental treatments for three dependent variables: *identification of best performing phase*, *identification of worst performing KPI* of a certain phase, *identification of the best performing KPI* of a certain phase.

A one-way analysis of variance (ANOVA) was also performed on the variables of interest across the three arms of the experiment, in order to elicit possible differences in the results compared to the aforementioned regression models. We ran ANOVA tests with the Bonferroni correction to adjust for the increased risk of a type-I error that arises when multiple statistical tests are being performed simultaneously.

Results

Our experimental results show significant differences in public managers' understanding of performance information across three different visual representations of equivalent data. Table 1

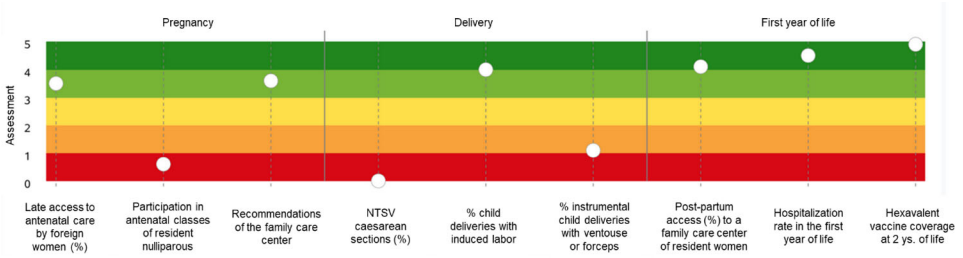

Figure 3. Pathway chart.

Table 1. Perceived understanding: average ratings, by treatment.

	Target chart ("Dartboard")	Radar chart	Pathway chart
Perceived understanding			
Clarity of the graph	3.25	3.01**	3.43^
Readability of the graph	3.16	2.85***	3.40***
Ease of recall	3.39	3.40	3.57*
vReadability of KPIs labels	3.67	3.84*	3.59
Clarity of the phases	3.11	2.99	3.39***

Significance of differences w.r.t. the target chart ("Dartboard").
^$p \leq 0.10$.
*$p \leq 0.05$.
**$p \leq 0.01$.
***$p \leq 0.001$.

reports average ratings for five measures of *perceived understanding*, by experimental condition. On average, self-reported perceptions of *clarity of the graph* were highest for hospital department heads presented with the pathway chart (3.43), followed by their counterparts in the dartboard arm of our experiment (3.25), and by participants who visualized the performance data as a radar chart (3.01). Data on self-reported *readability of the graph* indicate even sharper differences: the average level of readability was 3.16 for the dartboard, compared to 2.85 and 3.40 for the radar and the pathway charts, respectively. *Ease of recall* turned out to be higher for the pathway chart (3.57) relative to both the dartboard (3.39) and the radar chart (3.40), which did not significantly differ from each other with respect to this dimension of perceived understanding. As for *readability of KPI labels*, the radar chart (3.84) outperformed both the dartboard (3.67) and the pathway chart (3.59). As expected, hospital department heads in the pathway arm of the experiment perceived greater *clarity of performance by phase* (3.39), relative to both their counterparts in the target chart condition (3.11) and those exposed to the radar chart (2.99). Our findings hold true when controlling for the covariates included in the OLS regression models reported in Appendix 1. Moreover, results from a series of ANOVA tests provide further support for our conclusions (see Appendix 3).

Table 2 reports the percentages of hospital department heads who correctly answered each of the three questions in our test of *objective understanding*, by experimental condition. As expected, the proportion of subjects who correctly identified the *phase of the clinical pathway associated with the best performance* was noticeably higher among participants presented with the pathway chart (73%) relative to their counterparts exposed to either the dartboard (53%) or the radar chart (53%). The percentage of department heads who correctly indicated the *KPI with the worst score in the pregnancy phase* was higher for those presented with either a dartboard (59%) or a pathway chart (58%) compared to their peers who had seen the same data as a radar chart (45%). Finally, the proportion of subjects who identified the *KPI with the highest score in the delivery phase* was greater among department heads presented with the pathway chart (67%) relative to participants who had seen the same data as a dartboard (54%) or a radar chart (43%). These results held true

Table 2. Objective understanding: percentage of correct answers, by treatment.

	Target chart ("Dartboard")	Radar chart	Pathway chart
Objective understanding			
Phase with best performance	53%	53%	73%***
Lowest KPI in the pregnancy phase	59%	45%***	58%
Highest KPI in the delivery phase	54%	43%*	67%***

Significance of differences w.r.t. the target chart ("Dartboard").
^$p \leq 0.10$.
*$p \leq 0.05$.
**$p \leq 0.01$.
***$p \leq 0.001$.

when controlling for the covariates included in the logistic regression models in Appendix 2. Results from a series of ANOVA tests were consistent with our conclusions (Appendix 4).

A series of tests for interactions among our variables of interest did not indicate any significant effects.

Our findings should be interpreted in light of several limitations that point toward future research avenues. On the one hand, our experimental design ensures high internal validity through the random assignment of participants to experimental scenarios, which eliminates the risk of "systematic differences over conditions in respondents' characteristics that could also cause the observed effect" (Shadish, Cook, and Campbell 2002, 55). On the other hand, however, our study faces some of the external validity threats that are common in experimental studies for which participation is voluntary. Future work is certainly needed that tests whether and to what extent our inference may extend beyond our specific experimental setting and hold true across different types of public employees, organizational environments, and operations. Concerning the latter, in line with our attempt to maximize realism using graphical formats that are commonly adopted in our partner organizations, scholars engaged in replication studies may consider experimenting with alternative visual representations of performance data that are actually used—or could be—within public organizations in a different category.

Discussion and conclusions

We conducted a randomized controlled trial with 902 public hospital department heads to test whether alternative graphical representations of equivalent performance information may be conducive to different levels of perceived clarity and understanding, on the one hand, and objective recall and understanding, on the other hand. Our results demonstrate that the choice between alternative visualization techniques, which public healthcare organizations tend to use interchangeably, significantly affects both perceived and objective understanding by hospital department heads. Although all three techniques used in the study to visualize the overall maternal care performance (target, radar, and pathway chart) were intuitive and useful for displaying multidimensional health care data, the target and the pathway chart proved to be better suited than the radar chart to convey complex performance information. These results suggest at least four broad considerations.

First, our findings suggest that public managers have a good sense of which graphic information formats work best for them. This is implied by the similarity in the patterns of results that we observed in our measurements of subjective understanding, on the one hand, and objective recall of performance information, on the other. The *pathway chart* was first along most dimensions, followed by the target chart (a.k.a. "*dartboard*") and the *radar chart*, in that order. Our analysis does not allow drawing final conclusions about the nature of the information processing mechanisms underlying this pattern of results. Although our experimental design is well suited to estimate the average treatment effects of alternative graphical representations of the same

performance data on subjective understanding and objective recall, it does not illuminate the underlying chain reaction leading from our manipulations to their outcomes (Imai, Tingley, and Yamamoto 2013). This paves the way for future research that uses more sophisticated experimental designs—and triangulates quantitative and qualitative evidence to gain a richer understanding of the how in addition to the what (Mele and Belardinelli 2019).

Second, color coding can ease understanding of performance information by public professionals. Using colors to communicate meanings, as is the case for both the dartboard and the pathway chart, seems to enhance not only perceived clarity and understanding but also objective recall and understanding. This evidence supports that colors communicate immediate meanings, thus reducing users' cognitive load and shoring up the processes of knowledge acquisition (Keller et al. 2006). In the cases of the dartboard and the pathway chart, colors seem to convey specific messages (Isett and Hicks 2018) and reinforce the information offered via content structure.

Third, content structure can amplify or inhibit absorbtion of secondary or tacit information. Using headers aimed at locating each KPI within a certain phase of the service, as well as organizing information in linear or circular schemes, affect readers' capacity to extract relevant secondary or tacit information, such as the different phases of the maternal pathway. Studies on the combination of several attributes and dimensions are inconclusive about the superiority of three- versus two-dimensional structures because the former may "impose a considerable amount of additional extraneous cognitive load on learners due to the fact that they are usually associated with increased interactivity and additional orientation demands" (Keller et al. 2006, p. 48). On the contrary, the success of the pathway chart seems to suggest that a well-balanced informational combination, in addition to some meaningful content (i.e., the headers labeling the different phases of the pathway), may enhance information retention and comprehension.

Lastly, information ordering can convey metaphors. Isett and Hicks (2018) stated that when designing visualization tools, one should "consider culturally appropriate metaphors and layout schemas" (p. 482). In the case presented, different graphical presentational choices reflect different metaphors. The radar-chart (also known as *spider-chart*) conveys the idea of escaping from poor performance. The dartboard proposes a tacit message to 'hit the bull's eye'. The pathway chart (as it relies on the image of the musical stave) carries the idea of the 'patients' music', as it shows the women's path in the maternal care services for one year; performance 'sounds good' when the 'notes' (i.e. the KPIs) are positioned in the top bands of the 'stave'. Proximity of this latter metaphor to the clinical pathways, where physicians work on a daily basis, is evident. This can partly explain the better results of the pathway chart compared to the other representations. Furthermore, both the dartboard and the pathway chart "hint at a positive allusion, by referring to recreational and artistic activities" (Nuti et al. 2018, p. 2261); due to the hedonic quality of the appearance, these two graphical formats may attract and engage the reader more than the radar-chart.

Implications for practice are evident. The efficacy of the target-chart (i.e., the dartboard) is confirmed by policy practice, considering the number of organizations and governments that have decided to adopt it over time: for instance, the 2017 Country Health Profiles by the European Commission, the New Zealand dashboard developed by the Health Quality & Safety Commission, and the dartboard adopted by the Latvia Expert Group on Health Systems Performance Assessment (Noto et al. 2019). This study seems also to suggest that, when analyzing clinical pathways, a process-oriented representation (e.g., the pathway chart) can outperform the dartboard; this means that no one-size-fits-all solution can effectively address all data visualization issues, and culturally appropriate metaphors should be considered case by case in order to convey relevant information.

In the era of big data, public organizations need to develop impactful strategies to turn large amounts of figures and metrics into meaningful performance information that can foster the management of people and processes. Indeed, government and international organizations are

devoting ever-increasing attention and resources to the collection of performance data. Often neglected, however, is the presentation of those data in a way that maximizes their usability by public managers. A growing body of evidence indicates that performance information use can vary greatly across different visualizations of equivalent data (Ballard 2020). Nevertheless, too many public organizations still tend to consider alternative formats as interchangeable, and the choice between different visualization techniques as an irrelevant factor. This may be problematic because, as our findings demonstrate, data visualization does matter, and alternative graphical representations of equivalent data can result in significant differences in both perceived and actual understanding by public managers. In particular, color-coding seems to help reinforce the message of good and bad performance, whereas a balanced combination of content may enhance the comprehension and information retention. Moreover, the use of metaphors and layout schemas that resonate with managers and employees may greatly enhance their understanding.

Notes on contributors

Nicola Belle is a Researcher at the Management and Health Laboratory, within the Institute of Management and the EMbeDS Department of the Sant'Anna School of Advanced Studies. His work focuses on human resource management and decision-making in the context of public and healthcare organizations.

Giorgio Giacomelli is a Lecturer with the Government, Health and Not for Profit (GHNP) Division at SDA Bocconi School of Management, and Affiliate Researcher at the Institute of Management of Sant'Anna School of Advanced Studies. His research focuses on topics centered on public services performance and public sector HRM.

Sabina Nuti is the Rector of the Sant'Anna School of Advanced Studies and Full Professor of Health Management at the Institute of Management. She is the founder of the Management and Health Laboratory of the Sant'Anna School. Her research activity is focused on performance evaluation systems, efficiency, and governance mechanisms in health care.

Milena Vainieri is Associate Professor at the Management and Health Laboratory, within the Institute of Management and the EMbeDS Department of the Sant'Anna School of Advanced Studies. Her work focuses on performance management and governance mechanisms in the healthcare sector.

ORCID

Milena Vainieri http://orcid.org/0000-0002-0914-4487

References

Allwood, D., Z. Hildon, and N. Black. 2013. "Clinicians' Views of Formats of Performance Comparisons." *Journal of Evaluation in Clinical Practice* 19(1):86–93. doi: 10.1111/j.1365-2753.2011.01777.x.

Arnaboldi, M., I. Lapsley, and I. Steccolini. 2015. "Performance Management in the Public Sector: The Ultimate Challenge." *Financial Accountability & Management* 31(1):1–22. doi: 10.1111/faam.12049.

Baekgaard, M., N. Belle, S. Serritzlew, M. Sicilia, and I. Steccolini. 2019. "Performance Information in Politics: How Framing, Format, and Rhetoric Matter to Politicians' Preferences." *Journal of Behavioral Public Administration* 2(2):2. doi: 10.30636/jbpa.22.67.

Ballard, A. 2020. "Promoting Performance Information Use through Data Visualization: Evidence from an Experiment." *Public Performance and Management Review* 43(1):109–128.

Belardinelli, P., N. Bellé, and P. Cantarelli. 2020. "The Impact of Bounded Subadditivity on Administrative Behaviour among Public and Private Workers." *Public Administration*. doi: 10.1111/padm.12710.

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.

Bresciani, S., and M. J. Eppler. 2009. "The Benefits of Synchronous Collaborative Information Visualization: Evidence from an Experimental Evaluation." *IEEE Transactions on Visualization and Computer Graphics* 15(6): 1073–80. doi: 10.1109/TVCG.2009.188.

Brunell, T. L., and A. Glazer. 2001. "Rational Response to Irrational Attitudes: The Level of the Gasoline Tax in the United States." *Journal of Policy Analysis and Management* 20(4):761–4. doi: 10.1002/pam.1029.

Cantarelli, P., N. Belle, and P. Belardinelli. 2020. "Behavioral Public HR: Experimental Evidence on Cognitive Biases and Debiasing Interventions." *Review of Public Personnel Administration* 40(1):56–81. doi: 10.1177/0734371X18778090.

Desmidt, S., and K. Meyfroodt. 2021. "How Does Public Disclosure of Performance Information Affect Politicians' Attitudes towards Effort Allocation? Evidence from a Survey Experiment." *Journal of Public Administration Research and Theory* 31(4).

Dickinson, H., M. Bismark, G. Phelps, and E. Loh. 2016. "Future of Medical Engagement." *Australian Health Review* 40(4):443–6.

Dowding, D., R. Randell, P. Gardner, G. Fitzpatrick, P. Dykes, J. Favela, S. Hamer, Z. Whitewood-Moores, N. Hardiker, E. Borycki, et al. 2015. "Dashboards for Improving Patient Care: Review of the Literature." *International Journal of Medical Informatics* 84(2):87–100. doi: 10.1016/j.ijmedinf.2014.10.001.

Edwards, A., G. Elwyn, J. Covey, E. Matthews, and R. Pill. 2001. "Presenting Risk Information a Review of the Effects of Framing and Other Manipulations on Patient Outcomes." *Journal of Health Communication* 6(1): 61–82.

Franco-Santos, M., M. Kennerley, P. Micheli, V. Martinez, S. Mason, B. Marr, D. Gray, and A. Neely. 2007. "Towards a Definition of a Business Performance Measurement System." *International Journal of Operations & Production Management* 27(8):784–801. doi: 10.1108/01443570710763778.

Fryer, K., J. Antony, and S. Ogden. 2009. "Performance Management in the Public Sector." *International Journal of Public Sector Management* 22(6):478–98. doi: 10.1108/09513550910982850.

Fuenzalida, J., G. G. V. Ryzin, and A. L. Olsen. 2021. "Are Managers Susceptible to Framing Effects? An Experimental Study of Professional Judgment of Performance Metrics." *International Public Management Journal* 24(3):314–6. doi: 10.1080/10967494.2020.1752338.

Grimmelikhuijsen, S., S. Jilke, A. L. Olsen, and L. Tummers. 2017. "Behavioral Public Administration: Combining Insights from Public Administration and Psychology." *Public Administration Review* 77(1):45–56. doi: 10.1111/puar.12609.

Grover, V., R. H. L. Chiang, T.-P. Liang, and D. Zhang. 2018. "Creating Strategic Business Value from Big Data Analytics: A Research Framework." *Journal of Management Information Systems* 35(2):388–423. doi: 10.1080/07421222.2018.1451951.

Hildon, Z., D. Allwood, and N. Black. 2012a. "Impact of Format and Content of Visual Display of Data on Comprehension, Choice and Preference: A Systematic Review." *International Journal for Quality in Health Care* 24(1):55–64. doi: 10.1093/intqhc/mzr072.

Hildon, Z., D. Allwood, and N. Black. 2012b. "Making Data More Meaningful: Patients' Views of the Format and Content of Quality Indicators Comparing Health Care Providers." *Patient Education and Counseling* 88(2): 298–304. doi: 10.1016/j.pec.2012.02.006.

Huang, W., P. Eades, and S.-H. Hong. 2009. "Measuring Effectiveness of Graph Visualizations: A Cognitive Load Perspective." *Information Visualization* 8(3):139–52. doi: 10.1057/ivs.2009.10.

Imai, K., D. Tingley, and T. Yamamoto. 2013. "Experimental Designs for Identifying Causal Mechanisms: *Experimental Designs for Identifying Causal Mechanisms*." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176(1):5–51. doi: 10.1111/j.1467-985X.2012.01032.x.

Isett, K. R., and D. M. Hicks. 2018. "Providing Public Servants What They Need: Revealing the "Unseen" through Data Visualization." *Public Administration Review* 78(3):479–85. doi: 10.1111/puar.12904.

James, O., A. L. Olsen, D. P. Moynihan, and G. G. Van Ryzin. 2020. *Behavioral Public Performance: How People Make Sense of Government Metrics*. Cambridge: Cambridge University Press.

Johnsen, Å. 2005. "What Does 25 Years of Experience Tell us about the State of Performance Measurement in Public Policy and Management?" *Public Money and Management* 25(June):37–41. doi: 10.1111/j.1467-9302.2005.00445.x.

Kahneman, D. 2002. "Maps of Bounded Rationality: A Perspective on Intuitive Judgement and Choice." *Nobel Prize Lecture* 8:351–401. doi: 10.1055/s-2003-44981.

Kahneman, D. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review* 93(5):1449–1475. doi: 10.1257/000282803322655392.

Keller, T., P. Gerjets, K. Scheiter, and B. Garsoffky. 2006. "Information Visualizations for Knowledge Acquisition: The Impact of Dimensionality and Color Coding." *Computers in Human Behavior* 22(1):43–65. doi: 10.1016/j.chb.2005.01.006.

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

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.

Lindquist, E. 2011a. "Grappling with Complex Policy Challenges: Exploring the Potential of Visualization for Analysis, Advising and Engagement." HC Coombs Policy Forum, pp. 1–18.

Lindquist, E. 2011b. "Surveying the World of Visualization." HC Coombs Policy Forum, pp. 1–23.

Matheus, R., M. Janssen, and D. Maheshwari. 2020. "Data Science Empowering the Public: Data-Driven Dashboards for Transparent and Accountable Decision-Making in Smart Cities." *Government Information Quarterly* 37(3):101284.

McNeil, B. J., S. G. Pauker, J. H. C. Sox, and A. Tversky. 1982. "On the Elicitation of Preferences for Alternative Therapies." *The New England Journal of Medicine* 306(21):1259–62.

Mele, V., and P. Belardinelli. 2019. "Mixed Methods in Public Administration Research: Selecting, Sequencing, and Connecting." *Journal of Public Administration Research and Theory* 29(2):334–47. doi: 10.1093/jopart/muy046.

Modell, S. 2012. "Strategy, Political Regulation and Management Control in the Public Sector: Institutional and Critical Perspectives." *Management Accounting Research* 23(4):278–95. doi: 10.1016/j.mar.2012.05.001.

Morales-Silva, D. M., C. S. McPherson, G. Pineda-Villavicencio, and R. Atchison. 2020. "Using Radar Plots for Performance Benchmarking at Patient and Hospital Levels Using an Australian Orthopaedics Dataset." *Health Informatics Journal* 26(3):2119–37. doi: 10.1177/1460458219895190.

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

Murante, A. M., S. Nuti, and D. Matarrese. 2015. *Report of the Maternity Pathway*. Pp. 1–79. Firenze: Edizioni Polistampa.

Noto, G., I. Corazza, K. Klavina, J. Lepiskone, and S. Nuti. 2019. "Health System Performance Assessment in Small Countries: The Case Study of Latvia." *International Journal of Health Planning and Management* 34(4):1408–22.

Nuti, S., G. Noto, F. Vola, and M. Vainieri. 2018. "Let's Play the Patients Music: A New Generation of Performance Measurement Systems in Healthcare." *Management Decision* 56(10):2252–72. doi: 10.1108/MD-09-2017-0907.

Nuti, S., C. Seghieri, and M. Vainieri. 2013. "Assessing the Effectiveness of a Performance Evaluation System in the Public Health Care Sector: Some Novel Evidence from the Tuscany Region Experience." *Journal of Management & Governance* 17(1):59–69. doi: 10.1007/s10997-012-9218-5.

Nuti, S., C. Seghieri, M. Vainieri, and S. Zett. 2012. "Assessment and Improvement of the Italian Healthcare System: First Evidence from a Pilot National Performance Evaluation System." *Journal of Healthcare Management/American College of Healthcare Executives* 57(3):182–99.

Nuti, S., F. Vola, A. Bonini, and M. Vainieri. 2016. "Making Governance Work in the Health Care Sector: Evidence from a 'Natural Experiment' in Italy." *Health Economics, Policy, and Law* 11(1):17–38. doi: 10.1017/S1744133115000067.

Olsen, A. L. 2013. "Leftmost-Digit-Bias in an Enumerated Public Sector? An Experiment on Citizens' Judgment of Performance Information." *Judgment and Decision Making* 8(3):7.

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

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

Olsen, A. L. 2017a. "Human Interest or Hard Numbers? Experiments on Citizens' Selection, Exposure, and Recall of Performance Information." *Public Administration Review* 77(3):408–20. doi: 10.1111/puar.12638.

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

Pollitt, C. 2013. "The Logics of Performance Management." *Evaluation* 19(4):346–63. doi: 10.1177/1356389013505040.

Radnor, Z. J., and D. Barnes. 2007. "Historical Analysis of Performance Measurement and Management in Operations Management." *International Journal of Productivity and Performance Management* 56(5/6):384–96. doi: 10.1108/17410400710757105.

Rahman, A. A., Y. B. Adamu, and P. Harun. 2017. "Review on Dashboard Application from Managerial Perspective." International Conference on Research and Innovation in Information Systems, ICRIIS. doi: 10.1109/ICRIIS.2017.8002461.

Ryzin, G. G. V. 2013. "An Experimental Test of the Expectancy-Disconfirmation Theory of Citizen Satisfaction." *Journal of Policy Analysis and Management* 32(3):597–614. doi: 10.1002/pam.21702.

Shadish, W., T. D. Cook, and D. T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

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

Slingsby, A., J. Dykes, J. Wood, and R. Radburn. 2014. "Designing an Exploratory Visual Interface to the Results of Citizen Surveys." *International Journal of Geographical Information Science* 28(10):2090–125. doi: 10.1080/13658816.2014.920845.

Smith, P. C. 2005. "Performance Measurement in Health Care: History, Challenges and Prospects." *Public Money & Management* 25(4):213–20. doi: 10.1080/09540962.2005.10600096.

Stadler, J. G., K. Donlon, J. D. Siewert, T. Franken, and N. E. Lewis. 2016. "Improving the Efficiency and Ease of Healthcare Analysis through Use of Data Visualization Dashboards." *Big Data* 4(2):129–35. doi: 10.1089/big.2015.0059.

Stafoggia, Massimo, Adele Lallo, Danilo Fusco, Anna P. Barone, Mariangela D'Ovidio, Chiara Sorge, and Carlo A. Perucci. 2011. "Spie Charts, Target Plots, and Radar Plots for Displaying Comparative Outcomes of Health Care." *Journal of Clinical Epidemiology* 64(7):770–8. doi: 10.1016/j.jclinepi.2010.10.009.

Thaler, R. H., C. R. Sunstein, and J. P. Balz. 2013. "Choice Architecture." In E. Shafir (Ed.), *The Behavioral Foundations of Public Policy*. Princeton, NJ: Princeton University Press.

Tufte, E. 1990. *Envisioning Information*. Cheshire, CT: Graphic Press.

Tversky, A., and D. Kahneman. 1974. "Heuristics and Biases: Judgement under Uncertainty." *Science (New York, NY)* 185(4157):1124–30.

Tversky, A., and D. Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science (New York, NY)* 211(4481):453–8.

Vainieri, M., F. Ferrè, G. Giacomelli, and S. Nuti. 2019. "Explaining Performance in Health Care: How and When Top Management Competencies Make the Difference." *Health Care Management Review* 44(4):306–17. doi: 10.1097/HMR.0000000000000164.

Van Dooren, W., G. Bouckaert, and J. Halligan. 2015. *Performance Management in the Public Sector*. Milton Park, Abingdon-on-Thames: Routledge.

Woelert, P. 2015. "The "Logic of Escalation" in Performance Measurement: An Analysis of the Dynamics of a Research Evaluation System." *Policy and Society* 34(1):75–85. doi: 10.1016/j.polsoc.2015.03.004.

Woods, M., and S. Grubnic. 2008. "Linking Comprehensive Performance Assessment to the Balanced Scorecard: Evidence from Hertfordshire County Council." *Financial Accountability & Management* 24(3):343–61. doi: 10.1111/j.1468-0408.2008.00456.x.

Zwijnenberg, N. C., M. Hendriks, D. M. J. Delnoij, A. J. E. De Veer, P. Spreeuwenberg, and C. Wagner. 2016. "Understanding and Using Quality Information for Quality Improvement: The Effect of Information Presentation." *International Journal for Quality in Health Care* 28(6):689–97. doi: 10.1093/intqhc/mzw092.