Public Management Review

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rpxm20

Performance rankings reduce cognitive processing of underlying performance information

Lisa Hohensinn, Jurgen Willems, Bert George & Steven Van de Walle

To cite this article: Lisa Hohensinn, Jurgen Willems, Bert George & Steven Van de Walle (13 Feb 2025): Performance rankings reduce cognitive processing of underlying performance information, Public Management Review, DOI: 10.1080/14719037.2025.2464761

To link to this article: https://doi.org/10.1080/14719037.2025.2464761

Published online: 13 Feb 2025.

PUBLIC MANAGEMENT REVIEW
https://doi.org/10.1080/14719037.2025.2464761

Performance rankings reduce cognitive processing of underlying performance information

Lisa Hohensinn[a], Jurgen Willems[a], Bert George[b] and Steven Van de Walle[c]

[a]Department of Management, WU Vienna University of Economics and Business, Vienna, Austria; [b]Department of Public and International Affairs, City University of Hong Kong, Hong Kong, SAR; [c]Public Governance Institute, KU Leuven, Leuven, Belgium

ABSTRACT
Performance information is often presented in a ranked format. Rankings aggregate a multitude of performance dimensions into an overall score. Simultaneously, rankings may constrain cognitive processing of performance information because they distract users' attention away from the information underlying the ranking calculation. We test this adverse effect using university performance rankings in an eye-tracking experiment based on 1,071 decisions from 153 student-participants. Results show that performance rankings reduce cognitive processing of the underlying performance information, demonstrating the existence of a substitution effect. This study contributes to theorizing about and testing the effectiveness of performance management practices in public management.

ARTICLE HISTORY Received 28 December 2023; Accepted 3 February 2025

KEYWORDS Performance rankings; performance management; experiment; eye-tracking

Introduction

Rankings have become a very popular method of presenting performance information on public services for accountability purposes and for supporting decision-making (Fowles, George Frederickson, and Koppell 2016; Meijer 2007; Van de Walle and Roberts 2008; Van de Walle and van Delft 2015). Rankings 'measure current or past performance of comparable service units against one another' (Hood 2007, 95). They exist when there is a 'large N' of units doing similar things in a decentralized way. Examples include hospital rankings, league tables for universities, or school rankings (see e.g. Dill and Soo 2005; Horta 2009).

On the one hand, rankings benefit information users by aggregating a multitude of performance dimensions into an overall score and enabling benchmarking between relevant entities, with a view of improving public service performance (Gerrish 2016). By their very nature, rankings are composites, and thus reductionist (Bevan and Hood 2006). Accordingly, a one-dimensional summary measure (Pidd

CONTACT Lisa Hohensinn lisa.hohensinn@wu.ac.at
Supplemental data for this article can be accessed at https://doi.org/10.1080/14719037.2025.2464761

2005, 483) reduces multiple performance dimensions into a single one, which aids the presentation and interpretation of performance information. Ranked performance information allows users to find order and maintain a sense of control in an environment characterized by an exponentially growing information supply (Bevan and Hood 2006; Van de Walle and Roberts 2008). Therefore, rankings offer a simplified depiction of reality which is particularly useful in terms of complex information and/or the high number of evaluated entities. Taken together, rankings provide information users with a simplified, general evaluation, which should make performance evaluations easier.

On the other hand, due to their aggregative nature, rankings may obscure the underlying performance information that constitute them. Users as a result may ignore this underlying performance information and focus solely on the overall relative position of public organizations on the ranking instead. As the public sector, in contrast to the private sector, lacks clear performance measures and is affected by goal ambiguity (Rainey 2009), a summary measure for performance contradicts the multidimensional nature of public service performance and the inherent pluralism of public values underlying what constitutes 'good' or 'bad' performance in the public sector (Fukumoto and Bozeman 2019). A summary measure may prevent information users from reflecting on the complex nature of public service performance, which further affects evaluation, learning, or public values assessment (see e.g. Huijbregts, George, and Bekkers 2022a). Accordingly, decision-making in the public sector is characterized by balancing different public values, which requires making these values explicit through measurement. Consequently, ranking public service performance might be problematic, and a strong focus on performance rankings might impact cognitive processing of information and subsequent decision-making.

Apart from being used by public managers and policymakers, rankings are designed to influence the choices users make between public services (Le Grand 2007; Tummers, Jilke, and Van de Walle 2014). Whereas policy studies have looked at the effect of rankings on users' choices (Koning and van der Wiel 2013), we know very little about how users actually read and use rankings. A recurring question in scholarship on performance information use is 'how [users of public services] make sense of all this data' (Olsen 2017a, 562). In this study, we test whether performance rankings influence cognitive processing of the underlying performance information.

Accordingly, we adopted an eye-tracking laboratory experiment, in which we asked 153 participants – in multiple evaluation decisions – to evaluate universities' performance and make a preference decision based upon it. The experiment consisted of seven unrelated and randomized trials ($n = 1{,}071$ nested decisions). For each trial, we experimentally varied the presence of explicit ranking information on universities' performance. Our main dependent variable is the total fixation duration on pieces of information, as a close proxy for cognitive processing of information (Bera, Soffer, and Parsons 2019; Just and Carpenter 1976). Such measure is highly relevant to answering our research question, given various theoretical and empirical insights on how variations in visual attention to pieces of information clarify the extent to which these pieces of information are cognitively processed to actually make a decision (Walker et al. 2020). In this study, we focus on the preference decision to assess public service performance. Concretely, we hypothesize and test whether providing (additional) ranking information has an impact on which pieces of information receive more or less attention in the overall evaluation of public service performance.

By answering this question, we complement current public administration theory and research, especially the field of behavioural public administration that focuses on 'the understanding of how performance information is used by citizens and public managers' (James et al. 2020, 5). More specifically, first, this article contributes to the emerging literature studying presentational effects on performance information use (see e.g. Baekgaard and Serritzlew 2020; George et al. 2020; Jacobsen, Snyder, and Saultz 2014; James and Van Ryzin 2019; Piotrowski, Grimmelikhuijsen, and Deat 2019). This literature has largely drawn on cognitive bias and social norms as theoretical mechanisms explaining why and how users respond to performance information on public services. We contribute to this literature by introducing the substitution effect as a theoretical mechanism explaining the impact of how performance information is presented. An information substitution effect can be defined as the (relative) change in attention towards overall ranking information and away from more concrete information on which the ranking is based. Studying presentational effects such as the substitution effect is particularly relevant because public managers do not generally have the autonomy to manipulate concrete content of performance information, but can influence whether or not to report performance data and in which specific way (see e.g. Christensen and James 2022).

Second, we assume, in line with earlier work on performance information use, that different users 'respond fundamentally differently to performance information' (Baekgaard and Serritzlew 2020, 154). Earlier work has, for instance, shown that factors such as cognitive ability (Baekgaard and Serritzlew 2020), whether or not one is engaged or unengaged (Piotrowski, Grimmelikhuijsen, and Deat 2019), or levels of trust in government (James and Van Ryzin 2017) affect how people read and interpret performance information. Against this background, we add in our theoretical considerations and experimental design users' preferred style of cognitive processing (Cools and Van den Broeck 2007). In line with earlier eye-tracking studies by Mawad et al. (2015) and Koć-Januchta et al. (2017), as well as assumptions based on cognitive fit theory (Huang et al. 2006), we introduce cognitive styles as moderating variables to explain variation in the substitution effect of performance rankings between participants. In doing so, we can validate as well as elaborate the theoretical interpretation of how performance data is processed in relation to personal characteristics.

Third, extant experimental public administration research studying how service users or policymakers use performance information tends to use split ballot surveys and vignettes (e.g. James and Moseley 2014; Olsen 2017a). While these approaches allow for a greater number of subjects to be surveyed, they typically centre around measuring behavioural intentions as dependent variables as opposed to actual behaviour. Moreover, these approaches often employ fictional scenarios and/or fictional performance information on public services, resulting in criticism of the realistic nature of the experiment (Aguinis and Bradley 2014). Our eye-tracking experiment allows us to measure actual human behaviour, namely visual attention – closely related with cognitive processing (Bera, Soffer, and Parsons 2019; Just and Carpenter 1976) – of subjects based on eye movement and fixation, as opposed to measuring intentions based on replies to surveys. We also rely on real performance information on a topic relevant to our experiment subjects, i.e. students rate universities. In doing so, we validate and elaborate earlier findings with an alternative design, taking external validity into account.

Theoretical background

Substitution effect of ranking information

Rankings as a one-dimensional summary measure (Pidd 2005) are a popular method to present public performance information (e.g. Dill and Soo 2005; Horta 2009). The information presentation helps users interpret complex performance information from a high number of evaluated entities. The other unities serve as reference points and assist service users when interpreting performance information (Olsen 2017a; Webeck and Nicholson-Crotty 2020). Consequently, rankings work as decision aids and support users in evaluating entities' performance without studying more detailed performance information by themselves.

This, however, means that performance information users might focus on the summary measure only without having a detailed look into the performance dimensions that matter most to them. Although rankings help users evaluate performance, the performance reductionism could also have unintended consequences (Dixon et al. 2013). Rankings may encourage ignoring the multi-dimensional nature of performance that is not offered in a ranked format. An extensive focus on rankings rather than underlying data may lead to a measure fixation (Smith 1995), whereby the ranking position becomes more important than the underlying constituent parts of it.

The (relative) change in attention towards overall ranking information and away from more concrete information on which the ranking is based is referred to as a *substitution effect*. Substitution effects have been observed in studies on performance information use. For instance, Olsen (2017b) questioned whether hard performance numbers could crowd out more episodic, narrative information about public service performance. In contrast, Willems, Waldner, and Ronquillo (2019) found that explicit information signals such as star ratings do not constitute substitute information but are merely consulted as complementary information. Hence, given the theoretical considerations and similar effects in other concrete settings, our main hypothesis is that the presence of ranking information substitutes for other performance information, and rankings thus reduce cognitive processing for that other available information.

Hypothesis 1: Performance rankings reduce cognitive processing of the underlying information on which the ranking is based.

In our empirical analysis, we test whether the availability of summarizing and additional ranking information reduces the time participants spend consulting all available performance information. However, and starting from this overall hypothesis, we elaborate our argumentation by distinguishing between two concrete informational elements of performance ranking information: the number and the order element (see Figure 1).

First, there is the concrete summarizing information of ranking information, i.e. the ranking number. This means that the available information on a single organization or service is simultaneously grouped in a single metric and compared with the set of other relevant organizations/services. In other words, the concrete ranking number (e.g. '1st' and '29th', or 'First' and 'Twenty-ninth') aggregates the various performance dimensions into a single metric and compares them with the pooled performance of the other objects of comparison. This information can – according to the logic of our main

GROUP A No ranking information		**GROUP B** Ranking information (number)			**GROUP C** Ranking information (number and order)		
	performance indicators			performance indicators			performance indicators
Institution A	———	2	Institution A	———	1	Institution C	———
Institution B	———	4	Institution B	———	2	Institution A	———
Institution C	———	1	Institution C	———	3	Institution E	———
Institution D	———	5	Institution D	———	4	Institution B	———
Institution E	———	3	Institution E	———	5	Institution D	———

Figure 1. Number and order element of ranking information.

hypothesis – influence the cognitive processing of the underlying performance dimensions.

Second, and on top of the ranking number, ranking information often, but not necessarily, also contains an order element, i.e. the ranking order. This means that information is visually structured according to the summarizing ranking number. Concretely, this means that the best performing object is reported on top of the list, and the other objects are reported underneath, based on their ranking number in the set of compared objects.

This distinction between the number and order element of ranking information is particularly relevant when studying cognitive processing and visual attention. Consequently, we elaborate Hypothesis 1 by focusing on the number element (Hypothesis 1a) and the order element (Hypothesis 1b). It is important to emphasize the asymmetric dependence of Hypothesis 1b on Hypothesis 1a, as ordering objects based on ranking number also requires the ranking numbers to be reported.

Hypothesis 1a: The number element of performance rankings reduces cognitive processing of the underlying information on which the ranking is based.

Hypothesis 1b: The order element of performance rankings reduces cognitive processing of the underlying information on which the ranking is based (in addition to the number element effect).

Cognitive styles as moderator

Apart from testing the direct effect of performance rankings – both number and order – we also identify whether this effect becomes weaker or stronger among specific subjects based on their cognitive style. Information users differ in how they prefer to process information. Cognitive psychologists label these different preferred ways of processing information as one's cognitive style (Armstrong, Cools, and Sadler-Smith 2012). Cognitive styles have long been an area of investigation within business and management studies (see the review of Armstrong, Cools, and Sadler-Smith 2012), and have also been an inherent part of public management studies focused on the adoption and acceptance of performance and financial management practices (e.g. George et al. 2018; Kroll 2014). We build on earlier work on the role of cognitive ability and cognitive biases on the use of performance information (Baekgaard and Serritzlew 2020), and assume that cognitive styles moderate the substitution effect of performance rankings.

Cognitive styles have mostly been used in the context of the adoption of specific technology or management tools (Armstrong, Cools, and Sadler-Smith 2012). Traditionally, cognitive styles have been operationalized as a unidimensional and bipolar construct (i.e. two opposites of the same scale) distinguishing between an analytical and an intuitive way of processing information (Hodgkinson and Sadler-Smith 2003). Studies then showed that people with a more intuitive way of processing information are more likely to adopt new technology or management tools than those with an analytical approach (Chakraborty, Jen-Hwa Hu, and Cui 2008; Kroll 2014; Lu, Yu, and Lu 2001). Sadler-Smith (2009) expanded this unidimensional and bipolar measurement and argued that multidimensional views are necessary to measure cognitive styles. The underlying argument is that a person's cognitive style cannot just be measured using two opposites of the same scale, but rather should be considered a multidimensional concept in which each style is measured using a set of items. Following these recommendations, we use the three-dimensional Cognitive Style Indicator (CoSI) model by Cools and Van den Broeck (2007), which has been validated in different contexts and different countries (e.g. Cools, Van den Broeck, and Bouckenooghe 2009; George et al. 2018; Knockaert et al. 2015), and distinguishes between a creating, knowing and planning cognitive style.

We build on this three-fold distinction to theorize how cognitive processing of performance rankings along with other available performance information (Hypothesis 1) is related to cognitive styles. Cognitive fit theory (Huang et al. 2006) suggests that the impact of data visualization tools – such as performance rankings – is contingent upon the extent to which the visualization fits the user's cognitive style. Different cognitive styles result in different preferences regarding visualization and, subsequently, in different task performance (Engin and Vetschera 2017). People scoring high on the creating style tend to follow their intuition and gut-feeling when processing information (Cools and Van den Broeck 2007). In other words, they are comfortable when being faced with uncertainty and freedom and tend to favour innovation and creativity when making decisions. This also implies that they are less inclined to conduct thorough analyses and research before making specific decisions but, rather, make fast decisions based on their initial thoughts (Cools, Van den Broeck, and Bouckenooghe 2009; Cools, Van Den Broeck, and Evans 2008; George et al. 2018; Knockaert et al. 2015). We expect that people who score high on the creating cognitive style are more prone to using intuitive heuristics with the aim of making fast decisions, and performance rankings fit well with their preferences due to the relatively straightforward interpretation and visualization of rankings (Engin and Vetschera 2017).

Hypothesis 2: The negative effect of performance rankings on cognitive processing of the underlying information is stronger when participants have a creating cognitive style.

People scoring high on the planning style tend to favour structure and organization (Cools and Van den Broeck 2007). Their approach to processing information relies on doing things systematically and as efficiently as possible. They are typically favourable towards strong time management and efficient decision-making and are prone to following structure as opposed to content. This also implies that they are sensitive towards the way information is presented to them, searching for structure before making decisions (Cools, Van den Broeck, and Bouckenooghe 2009; Cools, Van Den Broeck, and Evans 2008; George et al. 2018; Knockaert et al. 2015). Because of their focus on structure, we expect that people who score high on the planning cognitive style are particularly prone to see performance rankings as a heuristic, a structural representation that allows them to make fast decisions. Indeed, the ordered and structured nature of such a ranking fits well with their own preferences concerning data visualization and interpretation (Engin and Vetschera 2017).

Hypothesis 3: The negative effect of performance rankings on cognitive processing of the underlying information is stronger when respondents have a planning cognitive style.

People scoring high on the knowing style tend to favour analysis and reasoning when processing information (Cools and Van den Broeck 2007). They seek to make informed decisions by conducting an extensive analysis of facts and figures as well as logical and rational arguments. This implies that information processing for people with a knowing style is in its very nature more reflective. They might take longer before coming to a decision because they favour analysis and reasoning based on data, as opposed to intuition or structure (Cools, Van den Broeck, and Bouckenooghe 2009; Cools, Van Den Broeck, and Evans 2008; George et al. 2018; Knockaert et al. 2015). We expect that people who score high on the knowing cognitive style are less triggered by heuristics and the need to make fast decisions due to their reflective and analytical nature. Indeed, the simplistic and aggregative nature of performance rankings does not fit well with their preference for in-depth analysis, data and understanding (Engin and Vetschera 2017).

Hypothesis 4: The negative effect of performance rankings on cognitive processing of the underlying information is weaker when respondents have a knowing cognitive style.

Methods

Experimental design

We use eye-tracking in a laboratory decision setting, which is a suitable method to study cognitive processing (Just and Carpenter 1976). We asked subjects to make performance evaluation decisions that are realistic and relevant for them. Eye-tracking has hitherto not been used very actively in organization research (Meißner and Oll 2019) nor in public administration scholarship (for a recent example see Walker et al. 2020). One notable exception in public administration is the work by Demaj (2017) and Demaj and Schedler (2014) on how performance information influences legislators' budgeting decisions.

The idea behind eye-tracking is that where people look at is what they are cognitively processing. When people consult information, their eye movement is rapidly changing between all the available information, altered with relatively longer fixations on particular pieces of information (in terms of milliseconds) (Glaholt and Reingold 2011; Rayner 1998). These fixations can be considered as a cognitive mechanism to process that particular information from all information available in the overall decision made (Glaholt and Reingold 2011). Quantifying visual fixation on pieces of information thus gives an insight in the extent that this information is considered in the overall decision of a participant. Hence, eye-tracking tools track the movement of subjects' eyes, and their fixations, as proxy for cognitive processing by the participant (Just and Carpenter 1976). Consequently, this study does not investigate the actual decision made, but the information focused on to make a decision, which is reporting a personal preference. Eye-tracking enables us thus to clarify the underlying cognitive processes while a decision is made (Glaholt and Reingold 2011).

We applied a between-subjects design with repeated trials. This means that participants were given seven information sheets in random order and for each of these sheets they were asked to make a decision. For each trial (i.e. an information sheet including a decision to make) they were randomly assigned to one of three groups. The difference between these three groups relates to Hypotheses 1a and 1b. In Group A, information for four performance dimensions was given (columns) for five objects to compare (rows: Objects were alphabetically ordered). In Group B, the same information was given with an extra column that reported rank number of the objects (Hypothesis 1a: Ranking numbers explicitly mentioned, but objects still alphabetically ordered). In Group C, the same information as in Group B was given (a summarizing rank number and four performance dimension columns for five objects), but with the objects ordered according to rank (Hypothesis 1b: Ranking numbers explicitly mentioned, and objects ordered according to this rank). Respondents were asked to choose each time their preferred object – with respect to best public service performance – based on all the available information. However, this actual decision outcome was not the main variable of interest, but the visual attention (fixation) to different pieces of information on the information sheets was. Moreover, we also make the critical (self-) reflection that our design has thus the concrete comparison of three situations where summarizing ranking information is (not) presented in addition to the underlying information on which the ranking information is based. Given our design, we do not test the additional difference of additional ranking information compared to any other additional related or unrelated information. We elaborate on this in our description of our complementary analyses. After the seven decisions, participants completed a questionnaire with items on cognitive styles and some demographics.

Participants

Students from the University of Hamburg were invited to participate in the laboratory experiment and were paid 10 euro for a 45-minute participation. In total, eight laboratory sessions were organized with a maximum of 30 participants per session. A standardized procedure was followed for each session. Students were informed before entering the laboratory that the experiment involved eye-tracking analysis, and once each student was assigned by chance to one of 30 experiment cabins, they were informed about the purpose of the experiment and the voluntary and anonymous nature of their participation. Additionally, all participants completed a consent form. Additional information on the sample can be found in the supplementary material.

Subsequently, the eye-tracking recording (Tobii Pro X2–60 eyetracker; attached to a 24-inch computer screen) was calibrated per participant, mainly to improve recoding quality. This calibration consisted of an initial basic task where respondents were asked to focus on five dots that appear in sequence on different places on the screen. With this procedure, which is standard for eye-tracking experiments in the laboratory, it can be evaluated, before continuing the recording for the actual experiment, whether the eye-tracking is functioning properly. This calibration, along with the seven decision trials and the subsequent survey questions, was administered with Z-tree (Fischbacher 1999; Fischbacher, Bendrick, and Schmid 2022).

In total, our analysis is based on 1,071 decisions of 153 participants. The average age was 26.17 (median: 25; min: 18; max: 46), and 57% were female. Moreover, 53% were bachelor students, 33% were master students, and 14% were enrolled in other programs (certification programs, PhD, etc.). The non-mandatory and open question on their study specialization shows a broad variety of backgrounds, including management, economics, psychology, sociology, law, political science, education sciences, engineering and informatics, and cultural and language sciences.

Framing, information sheets, and decision

Given the fact that we relied on students as experiment participants, we applied a framing and decision that is relevant in a student context. In other words, we decided for a performance evaluation task that relates to a public service in which students are an important stakeholder group. As a result, for seven countries (i.e. the seven trials) we listed five universities in that country along with performance data on research, teaching, funding, and internationalization. We asked them to select the university that they would prefer most as a potential collaboration partner of the University of Hamburg for exchange programs. Such task is relevant and realistic for the targeted population of our experiment. Moreover, the information provided for each country and university was real-life performance information from the Shanghai University Ranking initiative (ShanghaiRanking Consultancy 2022), per country (i.e. per trial) all respondents in all three groups received the exact same performance information. The only variation between the groups was the reporting of additional summarizing information (ranking number) (Group B compared to Group A), and ordering the university according to ranking number, instead of alphabetically (Group C compared to Group B).

Two additional design features are important to discuss in relation to the reliability and internal validity of our experiment. First, as earlier research with students on how they use university performance information found a small bias for one's own chosen university due to motivated reasoning (Christensen, 2018), we asked students from one university (University of Hamburg) to choose for seven other countries the university – out of a selection of five universities in that country – with which their own university should start collaborations for teaching and research. Second, people mainly use ranking information, report card metrics, or composite information when they do not already have access to or knowledge of other information regarding the topic (Kogan, Lavertu, and Peskowitz 2016; Lavertu 2016). Since we can assume that no matter what information is presented, our subjects would select Oxford or Cambridge in the UK, or Harvard and other Ivy League universities in the US, we have chosen universities in lesser-known countries, precisely because in such a context performance information plays an important role: no prior information available, faced with a complex reality. For the same reason, in some of the countries in the experiment, we did not include some highly ranked universities located in well-known cities.

Variables

Dependent variable

To analyse cognitive processing of information, we used total fixation duration (TFD) to sets of areas of interest (AOIs) in the information sheets. A fixation is a narrowly focused visual attention on a specific piece of information for a certain amount of time. These visual fixations are argued to be closely related to micro-level cognitive processing of that particular piece of information (Orquin and Mueller Loose 2013), and can be understood as a way in which people directly consult information while directly incorporating it in a broader evaluation of all available information to make a final decision. With respect to defining a fixation, we applied the Tobii Studio default settings: I-VT filter (velocity threshold of 30 degrees per second), merged adjacent fixations of a maximum time between fixations of 75 milliseconds, and maximum angle between fixations 0.5 degrees; average eye selection; and a minimum fixation duration of 60 milliseconds. Hence, we follow the suggestion of Orquin, Bagger, and Mueller Loose (2018) to apply a not too long of a minimum fixation duration to have sufficient data quality, which is also appropriate for our specific setting as the performance data we reported are two- or three-digit numbers, and not lengthy words or large numbers. Part A in Figure 2 gives a heat map reporting the accumulated total fixation duration for an information sheet for all participants (Example country China).

In the information sheets, we have marked areas of interest as depicted in Part B of Figure 2. Based on the recommendations of Orquin, Ashby, and Clarke (2016), the information sheets were designed to allow significant blank space between each piece of information to avoid overlap in fixation distribution and thus reduce the risk of false positive fixation registrations. As a result, for each of the AOIs, a total fixation duration (in seconds) is derived from the eye-tracking data. The sum of the total fixation durations on all AOIs without ranking related information is used as the total cognitive processing of underlying performance data. This overall sum included fixation on the concrete numbers for research, teaching, third-party funding, and internationalization. This sum is relevant for our main hypothesis testing, as we test whether the absence or presence of ranking information (number and order) influences cognitive processing on the underlying performance dimensions. This is measured in seconds that respondents have been focusing on pieces of information. As a result, our figures and tables provide values and estimates that report (differences in) cognitive processing time.

Independent variables

For each of the seven countries (trials in random order), respondents were randomly allocated to one of three groups. All three groups received the same basic performance information (in four columns) about five universities (rows). In Group A, the five universities were ordered alphabetically, and no ranking information was given. In Group B, the five universities were ordered alphabetically, and the ranking numbers were given in an additional column, left from the initial four columns with a: Heatmap Information

b: Areas of Interest

Figure 2. One out of seven decision sheets presented to participants.

performance data. In Group C, the ranking information was also given, but the universities were ordered consistently with the ranking numbers. In sum, our main independent variable is a categorical variable, with categories: 'Group A: Alphabetic, no ranking information (reference category)', 'Group B: Alphabetic, with ranking information', and 'Group C: Ranked, based on and with ranking information'.

Moderator variables

We measured cognitive styles with the three-dimensional construct of Cools and Van den Broeck (2007), and per cognitive style we averaged item scores. Items were measured with a 7-point Likert scale with labels 'strongly disagree' to 'strongly agree', and numeric labels were added ranging from '−3' to '+3'. Numeric labels centred on 0 are consistent with the assumption of a discrete scale variable, and it makes interpretation of our regression analyses more straightforward (see next section). Each dimension has convincing internal consistency: creating (7 items; Cronbach's alpha = 0.81); planning (4 items; Cronbach's alpha = 0.82); and knowing (7 items; Cronbach's alpha = 0.88).

Results

Main hypotheses

Figure 3 reports, aggregated over all seven decision cases, the mean values and 95% confidence intervals of the total fixation duration (TFD) per treatment group. These aggregated results show that, from an overall perspective, underlying performance information receives significantly less attention when additional ranking information is provided.

Table 1 reports three multi-level regression models where total fixation duration (TFD) to non-ranking information is each time the dependent variable, and where the seven decisions are nested within respondents. Hence, these analyses consider and control for multiple decisions made by 153 participants. It also allows to assess the extent to which each participant's cognitive processes are similar across the seven decisions made in this experiment. Across all three models, the intra-class correlations (ICC) are relatively high as they range between 0.55 and 0.58. This means that a substantial proportion of the observed cognitive processing is situated at the participant level.

In Model 1 only the treatment effects are reported, while in Model 2 the main effect between the three types of information sheets is tested, controlled for the data structure (random effects for nested data structure) and the seven different decisions (fixed effects for the order of the decision and for the country for which the decision was made). The intercept is 5.23 ($p < 0.001$), meaning that for a decision in the reference category, i.e. Group A (without ranking information), on average 5.23 seconds of total visual fixations occurred to any of the non-ranking information. It must be noted that this is the total fixation duration, which is the detailed and focused processing of specific performance numbers, and not the overall time a complete information sheet was consulted. For Groups B and C, where ranking information was added that summarizes the performance data, the fixation duration on the performance data was about 20% shorter (the coefficient for Group B = −1.12, $p < 0.001$; and the coefficient for Group C = −0.98; $p < 0.001$). These are reductions in cognitive processing of the underlying performance information as a result of ranking information of respectively 21.41% (1.12/5.23) and 18.74% (0.98/5.23). Consequently, Hypothesis 1a is supported, meaning that the additional reporting of ranking number draws away

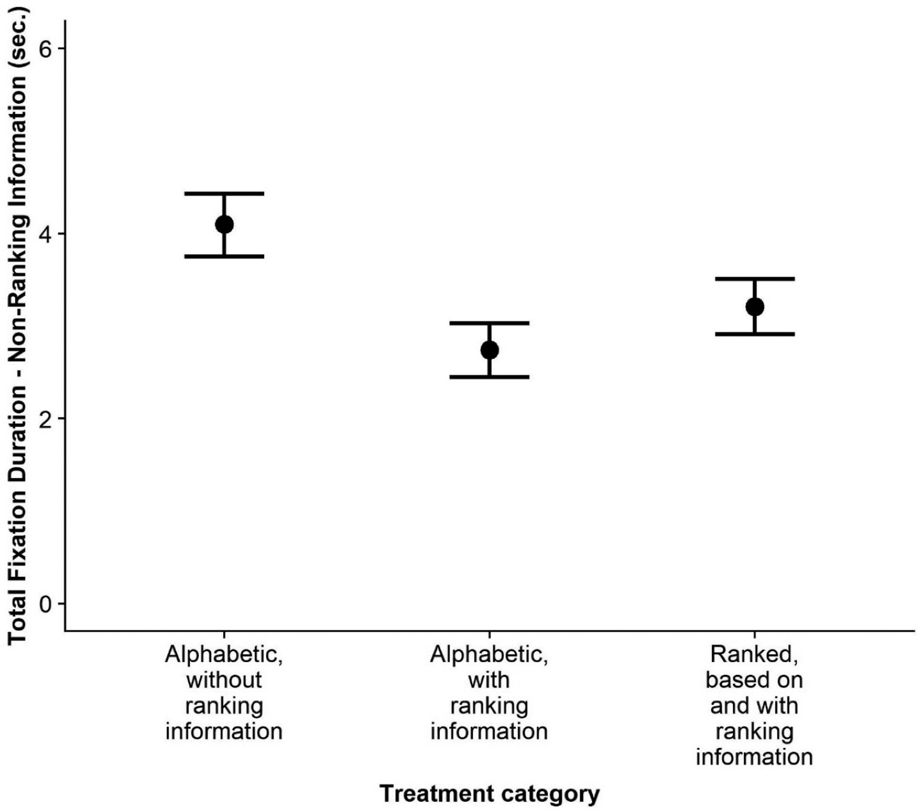

Figure 3. Group comparison for total fixation duration to all non-ranking information. **Note**: Group averages (total fixation duration in seconds) and confidence intervals based on 95% percentile-range are reported.

attention of the underlying information. There is no additional effect of structuring the information according to rank number. This means Hypothesis 1b is not supported. This can be derived from the confidence intervals in all four models and is depicted in Figure 3 (Kruskal–Wallis $\chi^2 = 39.213$, df = 2, $p < 0.001$). At this point, it has to be noted that complementary analyses, which are reported later in this article, have shown that respondents in all three settings spent similar amounts of time on all the information given to them, regardless of whether and how ranking information was given, but that the relative attention pattern to the different pieces of available information has changed.

Model 4 reports the same analysis but also includes the main effects of cognitive styles. As cognitive styles were rated on a scale from −3 to + 3, with 0 in the middle, the intercept of Model 3 reports the average time for a reference category decision (Group A: no ranking information), for a person that scores on the middle of all three cognitive style scales. However, no significant main effects exist, and the results of Model 3 are consistent with the results of Models 1 and 2.

Model 3 also includes the interaction effects of the decision treatment groups and the cognitive styles to test Hypotheses 2 to 4. Our results suggest that participants with a stronger creating cognitive style overall spend more processing time on non-ranking

Table 1. Explaining total fixation duration to non-ranking information.

	Model 1			Model 2			Model 3			Model 4		
	Estimates	*CI*	*P*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Group A (Intercept): Alphabetic, no ranking information (ref. cat.)	4.00	3.58–4.43	**<0.001**	5.23	4.64–5.83	**<0.001**	5.38	4.48–6.29	**<0.001**	5.51	4.55–6.48	**<0.001**
Group B: Alphabetic, with ranking information	−1.14	−1.46 – −0.81	**<0.001**	−1.12	−1.43 – −0.81	**<0.001**	−1.12	−1.43 – −0.81	**<0.001**	−1.08	−1.72 – −0.44	**0.001**
Group C: Ranked, based on and with ranking information	−0.96	−1.28 – −0.64	**<0.001**	−0.98	−1.29 – −0.67	**<0.001**	−0.98	−1.29 – −0.67	**<0.001**	−1.41	−2.03 – −0.80	**<0.001**
Cog. Style: Creating							−0.28	−0.71–0.15	0.194	−0.44	−0.91–0.04	0.071
Cog. Style: Knowing							0.03	−0.31–0.38	0.857	0.00	−0.38–0.38	0.984
Cog. Style: Planning							0.07	−0.31–0.45	0.711	0.10	−0.31–0.52	0.626
Interaction effects												
Group B × Creating										−0.02	−0.38–0.33	0.892
Group C × Creating										0.44	0.11–0.78	**0.010**
Group B × Knowing										0.12	−0.16–0.40	0.398
Group C × Knowing										−0.03	−0.30–0.24	0.84
Group B × Planning										−0.11	−0.42–0.19	0.472
Group C × Planning										0.01	−0.29–0.31	0.946
Control variables												
Decision sequence (random trials)				−0.28	−0.34 – −0.22	**<0.001**	−0.28	−0.34 – −0.22	**<0.001**	−0.28	−0.34 – −0.22	**<0.001**
Decision: Belgium				0.03	−0.42–0.48	0.89	0.03	−0.42–0.48	0.89	0.02	−0.43–0.47	0.919
Decision: Canada				−0.11	−0.56–0.34	0.627	−0.11	−0.56–0.34	0.627	−0.10	−0.55–0.35	0.669
Decision: China				−0.36	−0.81–0.09	0.114	−0.36	−0.81–0.09	0.114	−0.33	−0.79–0.12	0.15
Decision: Hong Kong				−0.57	−1.02 – −0.12	**0.014**	−0.57	−1.02 – −0.12	**0.014**	−0.52	−0.97 – −0.07	**0.025**
Decision: Netherlands				−0.04	−0.48–0.41	0.867	−0.04	−0.49–0.41	0.867	−0.03	−0.47–0.42	0.905
Decision: South Korea				0.28	−0.16–0.73	0.212	0.28	−0.16–0.73	0.212	0.29	−0.15–0.74	0.195
Random effects												
σ^2	4.27			3.9			3.9			3.88		
τ_{00}	5.17 $_{ID}$			5.23 $_{ID}$			5.27 $_{ID}$			5.30 $_{ID}$		
ICC	0.55			0.57			0.57			0.58		
N	153 $_{ID}$			153 $_{ID}$			153 $_{ID}$			153 $_{ID}$		
Observations	1,071			1,071			1,071			1,071		
Marginal R^2/Conditional R^2	0.025/0.559			0.060/0.598			0.067/0.603			0.071/0.607		

information (B = 0.44; $p = 0.010$), but only in the Group C, i.e. where ranking numbers are given, and information is structured according to these ranking numbers. This means that people with a high score for the creating cognitive style paid relatively more attention to underlying performance information when universities were ordered according to ranking information. As a result, partially contrasting support is found for Hypothesis 2, and no support for Hypotheses 3 and 4 as the other interaction terms have no significant coefficients. This means that the hypotheses derived from cognitive fit theory are not supported by our experiment. Here, it is important to point out that cognitive styles are an individual level variable, and that our laboratory design has 153 respondents at the individual level, with a relatively high ICC, meaning that individual factors rather than specific case or scenario factors determine variations. Hence, for these hypotheses our power is relatively low, at least compared to many other non-laboratory studies in the field of behavioural public administration. In contrast, as this is a laboratory study, we can for now conclude that such effects are either non-existing or not substantially large to be uncovered, even not in a controlled environment of a laboratory setting.

Finally, decision sequence, as a control variable, shows shorter fixation durations in later decisions ($\beta = -0.28$; $p < 0.001$), which is in line with previous research on eye-tracking studies that show that a learning effect exists, and that people need less processing time for every additional trial (Bagger et al. 2013; Orquin, Bagger, and Mueller Loose 2013). Moreover, students paid overall less attention to the decision to take about Hong Kong Universities ($\beta = -0.57$; $p < 0.014$).

Complementary analyses

To provide a sounder interpretation for the effects (not) confirmed, we perform a set of post-hoc complementary analyses that are reported in the Online Supplementary Materials. As these analyses are post hoc, caution is warranted about the interpretation of probability of effects. Nevertheless, results support exploration and postulation of new theoretical propositions. This is particularly valuable in the context of this study, given the new insights that eye-tracking data can give in performance evaluations and public administration research in general.

Complementary Analysis A analyzes the attention to all information – and not only the underlying ranking information – across all three scenarios. This shows that respondents attribute a more or less similar amount of information to any information sheet, regardless of whether and how ranking information is given. Hence, this complements our main findings, as the substitution effect is thus relative where a relatively stable amount of attention is redistributed over all available information. From this perspective, a critical note has to be made, and research needs to further validate to what extent this observed relative substitution is related to the specific ranking information, or the fact that additional information as such (and not necessarily ranking information) was given.

However, Complementary Analysis B decomposes the main findings by looking how relative attention for sub-parts of the information sheet changes as a result of the different scenarios. The overall substitution effect found from the main analysis is visible for the university performance dimensions (1) 'research', (2) 'teaching', and (3) 'funding', but not 'international'.

Moreover, while hypothesis 1b was not supported, Complementary Analysis C provides interesting insights related to the underlying reasoning for that hypothesis. In the alphabetic scenario without ranking information, and in the scenario with the order element of ranking information, the top two rows of information are relatively consulted more in depth, compared to the scenario with ranking information but not ordered according to the ranking. Moreover, rows 3, 4 and 5 are consulted relatively less when any kind of ranking information is given. This suggests that without ranking information, top rows are consulted more, which is in line with several findings of other eye-tracking studies. However, when ranking numbers are given (but not ordered accordingly), top rows do not receive this standard higher attention. When ranking information consists of number as well as order, top rows (again) do receive most attention. Hence, while extra information (in general, and for ranking information in the case of this study) can elicit an overall, but relative substitution effect (i.e. the same amount of attention is allocated differently), the specific ranking information (number or order) does seem to elicit different attention patterns.

Discussion

Do performance rankings influence cognitive processing of the underlying performance information? This research question lies at the heart of our study and is embedded in three observations: (1) decades of public management reform aimed at measuring and comparing the performance of public organizations across the globe (e.g. Hood 2012), (2) critical reflections towards performance measurement based on the negative effects surrounding performance rankings in modern-day governance (e.g. Diefenbach 2009), and (3) potential cognitive biases underlying performance measurement in public administration (e.g. James et al. 2020). Whereas previous research has by large been observational (see overview of Gerrish 2016) or, when experimental, mostly used survey experiments to measure attitudes and perceptions (see overview of Battaglio et al. 2019), we conducted an eye-tracking experiment, a method which has been given only scarce attention in public administration scholarship. Our findings indicate that rankings do indeed affect cognitive processing among our participants.

Specifically, our analysis shows that adding ranking information moves attention away from other performance information indicating the existence of a substitution effect. This is a partial crowding-out – or substitution – effect, as the overall processing time of all available information did not change as a result of available ranking information. Moreover, ranking information reported and applied by ordering units according to this rank results in relatively more attention for the top ranked units.

Implications for theory

A growing body of literature in behavioural public administration has focused on explaining why and how users respond to performance information on public services. A part of this literature has identified or confirmed cognitive bias and social norms as theoretical mechanisms. In contrast, also 'rational or smart heuristics' might be at play, for example when the ranking data is a fair and good representation of underlying data on which the overall ranking is based. In such case, a substitution effect can be considered as a rational time- and effort-saving cognitive mechanism to make sense of both underlying as well as summarizing ranking data. Hence, regardless of the theoretical perspective taken on this, and regardless of a rational reference framework that might be relevant and against which potential biases are evaluated (Willems, Damgaard, and Van Dooren 2019), our results help better understanding what the cognitive process is at the origin of these decision heuristics.

First, our findings indicate that additional information is considered, but additional information did not lead to additional attention or cognitive processing. This finding contributes to previous research by introducing the substitution effect as a theoretical mechanism explaining the impact of how performance information is presented. On the one hand, if the summarized information is of high quality and well-constructed, it may be questioned whether additional information is indeed relevant for a 'good' decision. Additional information could make decision-makers feel overwhelmed and hinder the decision-making process, as detailed information has to be additionally processed. On the other hand, if the other information was crucial or the only relevant element for an optimal decision, lower levels of cognitive processing of that information might explain why some people are making less optimal decisions. However, like in many decisions made by service users, policymakers, and public servants, no optimal decision had to be made in our setting, but only a personal preference was probed for.

Nevertheless, this finding is particularly relevant in a public sector setting, because public organizations, in contrast to private organizations, have multiple objectives and lack distinct cross-organizational performance measures (Rainey 2009). As public service performance is thus multidimensional, a ranking of public services that is based on a one-dimensional summary measure hardly catches the nature of public services. If individuals do not cognitively process additional information such as various performance measures, the question arises whether rankings are useful for conveying the multidimensionality of public service performance. Our results thus add to the discussion of how additional and different types of information alter the cognitive processing pattern.

Second, our finding that top-ranked organizations receive relatively more attention when visually ordered according to this rank contributes to the literature on how decision aids work in public performance reporting (see e.g. Christensen and James 2022). Concretely, our results show that rankings help put attention on best-performing units. If this is the purpose, then rankings can indeed be an effective way to report the main take-aways of a complex performance evaluation and comparison. Again, this finding has important implications for the public sector context. As public performance rankings are aggregations of multi-dimensional performance, this finding can result in levelling out – or even obscuring – performance dimensions that are, for example, only relevant for particular groups in society or in particular settings. Furthermore, an organization might perform well on one performance dimension and not so well on others (Holm 2018; Moynihan 2008). Aggregating the scores in a single composite ranking score ignores the goal multiplicity of public organizations, such as universities' different functions (Williams and de Rassenfosse 2016). In addition, comparing public entities suffers from conceptual problems, meaning a common understanding of what it means for a public entity to perform well is needed (see e.g. Van de Walle 2008). Against this background, our results contribute to the usability of rankings, depending on the purpose of the public performance reporting (Van de Walle and van Delft 2015).

Finally, our experiment fails to substantiate claims made by cognitive-fit theory about the role of cognitive styles in information processing. Although some evidence was uncovered about the role of a creating cognitive style, this evidence ran counter to our hypothesis. Consequently, our study does not demonstrate that certain cognitive styles are more susceptible to the substitution effect than others based on their preferred way of information processing. In contrast, previous – mostly observational – research has found that cognitive styles are associated with several performance management practices and performance information use (e.g. Kroll 2014). However, these studies typically identify direct associations between cognitive styles and performance management concepts based on survey responses from public managers. Anyway, the controlled context of our experiment could influence how cognitive styles materialize in behaviour, and students might differ from public managers in the manifestation of cognitive styles in behaviour. To broaden our understanding about the role of the research context, more theorizing in terms of the role cognitive styles play (or not) in how people respond to performance rankings is needed.

Implications for further research

We especially encourage future replications of our experiment aimed at empirical generalization and conceptual extension (Walker, James, and Brewer 2017). Replications 'can help to establish the external validity of knowledge and thus the ability to generalize a study's findings more broadly' (Walker et al. 2019, 609), and can be particularly helpful in tackling some of the study's limitations. While our findings clearly demonstrate the existence of a substitution effect, there are many design choices embedded in our experiment that future research could build on.

First, our sample consists of students from one university. Empirical generalization could take shape in two forms. Scholars could use a similar sample of students to replicate our experiment and identify whether our findings hold or focus on better representing the broader population or specific populations like service users, public managers, and policymakers. Either way, empirical generalization through a different sample using the same design can help assess the external validity of the existence of a substitution effect due to performance rankings.

Second, apart from exact replication using different samples, future research can implement several conceptual extensions. We highlight three important extensions: a different public service, a different ranking design, and a different set of moderators. While we focused on performance rankings in relation to universities, which are prevalent and receive loads of media attention, other public services are also often ranked, for example, 'best places to live' or 'best hospitals'. Future research could identify whether the substitution effect is similarly prevalent across policy domains or whether between-domain differences pop up, for instance, based on the salience of the domain (e.g. Huijbregts, George, and Bekkers 2022b). In terms of ranking design, we used a very straightforward design through a table including performance information. A fruitful research avenue remains of whether other designs, for instance, using a certain amount of stars as an indicator of rank (e.g. Willems, Waldner, and Ronquillo 2019) have a similar substitution effect or this effect might be mitigated through specific designs. Moreover, for now we do find a substitution effect based on the additional information provided, which was ranking information for the case of this study. Besides, our supplementary analyses suggest that additional ranking information (numbers versus order) also leads to other attention patterns, within the more-or-less stable amount of attention that respondents give to these information sheets. Further research is needed to also test and disentangle substitution effects in general, and specifically for ranking information. For example, which (ranking) information would potentially also reduce or increase overall information search patterns (and not only relative allocation of attention over all available information), and/or what other types of information would have similar or dissimilar effects.

Finally, we focused and theorized on cognitive styles as potential psychological characteristics of participants that could influence the substitution effect but did not uncover significant interactions. Further research could assess whether there might be additional moderators that could mitigate the substitution effect either at the individual level, e.g. other psychological characteristics such as the big five personality characteristics (Aarøe et al. 2021), or at the team/organizational level, e.g. procedural rationality of the decision-making process (George and Desmidt 2018).

Conclusion

Performance rankings remain extremely popular tools in public governance and management. Despite their popularity, however, these rankings are criticized for the often narrow, blinkered approach in which these are used and interpreted. We uncovered that rankings move attention away from other performance information, which seemingly supports much of this criticism and urges scholars and practitioners to reflect on whether, when and how performance rankings can be useful for learning and informed decision-making. We also hope to encourage research that uses experimental methods – such as eye-tracking – that go beyond measuring participants attitudes and perceptions towards actual behaviour, as these methods remain underused in public administration scholarship. For now, we can conclude that among the already impressive list of potential perverse effects related to performance rankings, we can add the danger of a substitution effect.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Aarøe, Lene, Martin Baekgaard, Julian Christensen, and Donald P. Moynihan. 2021. "Personality and Public Administration: Policymaker Tolerance of Administrative Burdens in Welfare Services." *Public Administration Review* 81 (4): 652–663. https://doi.org/10.1111/puar.13381.

Aguinis, Herman, and Kyle J. Bradley. 2014. "Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies." *Organizational Research Methods* 17 (4): 351–371. https://doi.org/10.1177/1094428114547952.

Armstrong, Steven J., Eva Cools, and Eugene Sadler-Smith. 2012. "Role of Cognitive Styles in Business and Management: Reviewing 40 Years of Research." *International Journal of Management Reviews* 14 (3): 238–262. https://doi.org/10.1111/j.1468-2370.2011.00315.x.

Baekgaard, Martin, and Søren Serritzlew. 2020. "Those Who Understand it Will Not Be Persuaded: A Performance Information Paradox." *International Public Management Journal* 23 (1): 138–160. https://doi.org/10.1080/10967494.2018.1461152.
Bagger, Martin P., Jacob L. Orquin, Susann Fiedler, and Simone Mueller Loose. 2013. "Capturing Learning Effects on Eye Movements in Repeated Measures Experiments - Forskning - Aarhus Universitet." *Lund, Sverige*. http://ecem2013.eye-movements.org.
Battaglio, R. Paul, Paolo Belardinelli, Nicola Bellé, and Paola Cantarelli. 2019. "Behavioral Public Administration Ad Fontes: A Synthesis of Research on Bounded Rationality, Cognitive Biases, and Nudging in Public Organizations." *Public Administration Review* 79 (3): 304–320. https://doi.org/10.1111/puar.12994.
Bera, Palash, Pnina Soffer, and Jeffrey Parsons. 2019. "Using Eye Tracking to Expose Cognitive Processes in Understanding Conceptual Models." *MIS Quarterly* 43 (4): 1105–1126. https://doi.org/10.25300/MISQ/2019/14163.
Bevan, Gwyn, and Christopher Hood. 2006. "What's Measured is What Matters: Targets and Gaming in the English Public Health Care System." *Public Administration* 84 (3): 517–538. https://doi.org/10.1111/j.1467-9299.2006.00600.x.
Chakraborty, Indranil, Paul Jen-Hwa Hu, and Dai Cui. 2008. "Examining the Effects of Cognitive Style in Individuals' Technology Use Decision Making." *Decision Support Systems* 45 (2): 228–241. https://doi.org/10.1016/j.dss.2007.02.003.
Christensen, J. 2018. "Biased, Not Blind: An Experimental Test of Self□Serving Biases in Service users' Evaluations of Performance Information." *Public Administration* 96 (3): 468–480.
Christensen, Julian, and Oliver James. 2022. "Reporting Multiple Dimensions of Public Service Performance: Information Order Effects on Citizens' Willingness to Use Services." *Public Management Review* 24 (1): 142–157. https://doi.org/10.1080/14719037.2020.1798708.
Cools, Eva, and Herman Van den Broeck. 2007. "Development and Validation of the Cognitive Style Indicator." *The Journal of Psychology* 141 (4): 359–387. https://doi.org/10.3200/JRLP.141.4.359-388.
Cools, Eva, Herman Van den Broeck, and Dave Bouckenooghe. 2009. "Cognitive Styles and Person–Environment Fit: Investigating the Consequences of Cognitive (Mis)fit." *European Journal of Work & Organizational Psychology* 18 (2): 167–198. https://doi.org/10.1080/13594320802295540.
Cools, Eva, Herman Van Den Broeck, and Carol Evans. 2008. "Cognitive Styles and Managerial Behaviour: A Qualitative Study." Edited by Martin Graff." *Education + Training* 50 (2): 103–114. https://doi.org/10.1108/00400910810862092.
Demaj, Labinot. 2017. "What Can Performance Information Do to Legislators? A Budget-Decision Experiment with Legislators." *Public Administration Review* 77 (3): 366–379. https://doi.org/10.1111/puar.12713.
Demaj, Labinot, and Kuno Schedler. 2014. "What Drives Legislators' Information Behavior? An Eye-Tracking Experiment with Legislators." Speyer, Germany. https://papers.ssrn.com/abstract=2494760.
Diefenbach, Thomas. 2009. "New Public Management in Public Sector Organizations: The Dark Sides of Managerialistic 'Enlightenment'." *Public Administration* 87 (4): 892–909. https://doi.org/10.1111/j.1467-9299.2009.01766.x.
Dill, David D., and Maarja Soo. 2005. "Academic Quality, League Tables, and Public Policy: A Cross-National Analysis of University Ranking Systems." *Higher Education* 49 (4): 495–533. https://doi.org/10.1007/s10734-004-1746-8.
Dixon, Ruth, Christiane Arndt, Manuel Mullers, Jarmo Vakkuri, Kristiina Engblom-Pelkkala, and Christopher Hood. 2013. "A Lever for Improvement or a Magnet for Blame? Press and Political Responses to International Educational Rankings in Four Eu Countries." *Public Administration* 91 (2): 484–505. https://doi.org/10.1111/padm.12013.
Engin, Ayşegül, and Rudolf Vetschera. 2017. "Information Representation in Decision Making: The Impact of Cognitive Style and Depletion Effects." *Decision Support Systems* 103 (November): 94–103. https://doi.org/10.1016/j.dss.2017.09.007.
Fischbacher, Urs. 1999. "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10 (2): 171–178. https://doi.org/10.1007/s10683-006-9159-4.
Fischbacher, Urs, Katharine Bendrick, and Stefan Schmid. 2022. "Z-Tree 3.5 - Tutorial and Reference Manual." https://www.ztree.uzh.ch/static/doc/manual_v3.pdf.
Fowles, Jacob, H. George Frederickson, and Jonathan G. S. Koppell. 2016. "University Rankings: Evidence and a Conceptual Framework." *Public Administration Review* 76 (5): 790–803. https://doi.org/10.1111/puar.12610.

Fukumoto, Eriko, and Barry Bozeman. 2019. "Public Values Theory: What is Missing?" *The American Review of Public Administration* 49 (6): 635–648. https://doi.org/10.1177/0275074018814244.

George, Bert, Martin Baekgaard, Adelien Decramer, Mieke Audenaert, and Stijn Goeminne. 2020. "Institutional Isomorphism, Negativity Bias and Performance Information Use by Politicians: A Survey Experiment." *Public Administration* 98 (1): 14–28. https://doi.org/10.1111/padm.12390.

George, Bert, and Sebastian Desmidt. 2018. "Strategic-Decision Quality in Public Organizations: An Information Processing Perspective." *Administration & Society* 50 (1): 131–156. https://doi.org/10.1177/0095399716647153.

George, Bert, Sebastian Desmidt, Eva Cools, and Anita Prinzie. 2018. "Cognitive Styles, User Acceptance and Commitment to Strategic Plans in Public Organizations: An Empirical Analysis." *Public Management Review* 20 (3): 340–359. https://doi.org/10.1080/14719037.2017.1285112.

Gerrish, Edited by. 2016. "The Impact of Performance Management on Performance in Public Organizations: A Meta-Analysis." *Public Administration Review* 76 (1): 48–66. https://doi.org/10.1111/puar.12433.

Glaholt, Mackenzie G., and Eyal M. Reingold. 2011. "Eye Movement Monitoring as a Process Tracing Methodology in Decision Making Research." *Journal of Neuroscience, Psychology, and Economics* 4 (2): 125–146. https://doi.org/10.1037/a0020692.

Hodgkinson, Gerard P., and Eugene Sadler-Smith. 2003. "Complex or Unitary? A Critique and Empirical Re-Assessment of the Allinson-Hayes Cognitive Style Index." *Journal of Occupational & Organizational Psychology* 76 (2): 243–268. https://doi.org/10.1348/096317903765913722.

Holm, Jakob Majlund. 2018. "Successful Problem Solvers? Managerial Performance Information Use to Improve Low Organizational Performance." *Journal of Public Administration Research & Theory* 28 (3): 303–320. https://doi.org/10.1093/jopart/muy017.

Hood, Christopher. 2007. "Public Service Management by Numbers: Why Does it Vary? Where Has it Come From? What are the Gaps and the Puzzles?" *Public Money and Management* 27 (2): 95–102. https://doi.org/10.1111/j.1467-9302.2007.00564.x.

Hood, Christopher. 2012. "Public Management by Numbers as a Performance-Enhancing Drug: Two Hypotheses." *Public Administration Review* 72 (s1): 85–92. https://doi.org/10.1111/j.1540-6210.2012.02634.x.

Horta, Hugo. 2009. "Global and National Prominent Universities: Internationalization, Competitiveness and the Role of the State." *Higher Education* 58 (3): 387–405. https://doi.org/10.1007/s10734-009-9201-5.

Huang, Zan, Hsinchun Chen, Fei Guo, Jennifer J. Xu, Soushan Wu, and Wun-Hwa Chen. 2006. "Expertise Visualization: An Implementation and Study Based on Cognitive Fit Theory." *Decision Support Systems* 42 (3): 1539–1557. https://doi.org/10.1016/j.dss.2006.01.006.

Huijbregts, Rowie, Bert George, and Victor Bekkers. 2022a. "Public Values Assessment as a Practice: Integration of Evidence and Research Agenda." *Public Management Review* 24 (6): 840–859. https://doi.org/10.1080/14719037.2020.1867227.

Huijbregts, Rowie, Bert George, and Victor Bekkers. 2022b. "Valuation Tools and Politicians' Willingness to Sell Public Real Estate: A Survey Experiment." *Public Management Review* 24 (6): 882–902. https://doi.org/10.1080/14719037.2021.1874496.

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

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

James, Oliver, and Gregg G. Van Ryzin. 2017. "Incredibly Good Performance: An Experimental Study of Source and Level Effects on the Credibility of Government." *The American Review of Public Administration* 47 (1): 23–35. https://doi.org/10.1177/0275074015580390.

James, Oliver, and Gregg G. Van Ryzin. 2019. "Rates and the Judgment of Government Performance." *Journal of Behavioral Public Administration* 2 (2): 1–10. https://doi.org/10.30636/jbpa.22.41.

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 University Press.

Just, Marcel Adam, and Patricia A Carpenter. 1976. "Eye Fixations and Cognitive Processes." *Cognitive Psychology* 8 (4): 441–480. https://doi.org/10.1016/0010-0285(76)90015-3.

Knockaert, Mirjam, Maw Der Foo, Truls Erikson, and Eva Cools. 2015. "Growth Intentions Among Research Scientists: A Cognitive Style Perspective." *Technovation* 38 (April): 64–74. https://doi.org/10.1016/j.technovation.2014.12.001.

Koć-Januchta, Marta, Tim Höffler, Gun-Brit Thoma, Helmut Prechtl, and Detlev Leutner. 2017. "Visualizers Versus Verbalizers: Effects of Cognitive Style on Learning with Texts and Pictures – an Eye-Tracking Study." *Computers in Human Behavior* 68 (March): 170–179. https://doi.org/10.1016/j.chb.2016.11.028.

Kogan, Vladimir, Stéphane Lavertu, and Zachary Peskowitz. 2016. "Performance Federalism and Local Democracy: Theory and Evidence from School Tax Referenda." *American Journal of Political Science* 60 (2): 418–435. https://doi.org/10.1111/ajps.12184.

Koning, Pierre, and Karen van der Wiel. 2013. "Ranking the Schools: How School-Quality Information Affects School Choice in the Netherlands." *Journal of the European Economic Association* 11 (2): 466–493. https://doi.org/10.1111/jeea.12005.

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

Lavertu, Stéphane. 2016. "We All Need Help: 'Big Data' and the Mismeasure of Public Administration." *Public Administration Review* 76 (6): 864–872. https://doi.org/10.1111/puar.12436.

Le Grand, Julian 2007. "The Other Invisible Hand: Delivering Public Services Through Choice and Competition." In *The Other Invisible Hand*, Princeton University Press. https://doi.org/10.1515/9781400828005.

Lu, Hsi-Peng, Huei-Ju Yu, and Simon S. K Lu. 2001. "The Effects of Cognitive Style and Model Type on DSS Acceptance: An Empirical Study." *European Journal of Operational Research* 131 (3): 649–663. https://doi.org/10.1016/S0377-2217(00)00107-7.

Mawad, Franco, Marcela Trías, Ana Giménez, Alejandro Maiche, and Gastón Ares. 2015. "Influence of Cognitive Style on Information Processing and Selection of Yogurt Labels: Insights from an Eye-Tracking Study." *Food Research International* 74 (August): 1–9. https://doi.org/10.1016/j.foodres.2015.04.023.

Meijer, Albert Jacob. 2007. "Publishing Public Performance Results on the Internet: Do Stakeholders Use the Internet to Hold Dutch Public Service Organizations to Account?" *Government Information Quarterly* 24 (1): 165–185. https://doi.org/10.1016/j.giq.2006.01.014.

Meißner, Martin, and Josua Oll. 2019. "The Promise of Eye-Tracking Methodology in Organizational Research: A Taxonomy, Review, and Future Avenues." *Organizational Research Methods* 22 (2): 590–617. https://doi.org/10.1177/1094428117744882.

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

Olsen, Asmus Leth. 2017a. "Compared to What? How Social and Historical Reference Points Affect Citizens' Performance Evaluations." *Journal of Public Administration Research & Theory* 27 (4): 562–580. https://doi.org/10.1093/jopart/mux023.

Olsen, Asmus Leth. 2017b. "Human Interest or Hard Numbers? Experiments on Citizens' Selection, Exposure, and Recall of Performance Information." *Public Administration Review* 77 (3): 408–420. https://doi.org/10.1111/puar.12638.

Orquin, Jacob L., Nathaniel J. S. Ashby, and Alasdair D. F. Clarke. 2016. "Areas of Interest as a Signal Detection Problem in Behavioral Eye-Tracking Research." *Journal of Behavioral Decision Making* 29 (2–3): 103–115. https://doi.org/10.1002/bdm.1867.

Orquin, Jacob L., Martin P. Bagger, and Simone Mueller Loose. 2013. "Learning Affects Top Down and Bottom Up Modulation of Eye Movements in Decision Making." *Judgment & Decision Making* 8 (6): 700–716. https://doi.org/10.1017/S1930297500004733.

Orquin, Jacob L., and Simone Mueller Loose. 2013. "Attention and Choice: A Review on Eye Movements in Decision Making." *Acta Psychologica* 144 (1): 190–206. https://doi.org/10.1016/j.actpsy.2013.06.003.

Orquin, J. L., and K. Holmqvist 2018. "Threats to the Validity of Eye-Movement Research in Psychology." *Behavior Research Methods* 50: 1645–1656.

Pidd, M. 2005. "Perversity in Public Service Performance Measurement." *International Journal of Productivity & Performance Management* 54 (5/6): 482–493. https://doi.org/10.1108/17410400510604601.

Piotrowski, Suzanne, Stephan Grimmelikhuijsen, and Felix Deat. 2019. "Numbers Over Narratives? How Government Message Strategies Affect Citizens' Attitudes." *Public Performance & Management Review* 42 (5): 1005–1028. https://doi.org/10.1080/15309576.2017.1400992.

Rainey, Hal G. 2009. *Understanding and Managing Public Organizations*, 4th edition. San Francisco, CA: John Wiley & Sons.

Rayner, Keith. 1998. "Eye Movements in Reading and Information Processing: 20 Years of Research." *Psychological Bulletin* 124 (3): 372–422. https://doi.org/10.1037/0033-2909.124.3.372.

Sadler-Smith, Eugene. 2009. "A Duplex Model of Cognitive Style." In *Perspectives on the Nature of Intellectual Styles*, edited by Robert J. Sternberg and Li-Fang Zhang, 3–28. New York: Springer Publishing Company.

Shanghai Ranking Consultancy. 2022. "Academic Ranking of World Universities." https://www.shanghairanking.com/rankings.

Smith, Peter. 1995. "On the Unintended Consequences of Publishing Performance Data in the Public Sector." *International Journal of Public Administration* 18 (2–3): 277–310. https://doi.org/10.1080/01900699508525011.

Tummers, Lars, Sebastian Jilke, and Steven Van de Walle. 2014. "Citizens in Charge? Reviewing the Background and Value of Introducing Choice and Competition in Public Services." In *Public Administration Reformation*, Routledge.

Van de Walle, Steven 2008. "Comparing the Performance of National Public Sectors: Conceptual Problems." *International Journal of Productivity & Performance Management* 57 (4): 329–338. https://doi.org/10.1108/17410400810867535.

Van de Walle, Steven, and Alasdair Roberts. 2008. "Publishing Performance Information: An Illusion of Control?" In *Performance Information in the Public Sector: How it is Used*, edited by Wouter Van Dooren and Steven Van de Walle, 211–226. Governance and Public Management Series. London: Palgrave Macmillan UK. https://doi.org/10.1007/978-1-137-10541-7_14.

Van de Walle, Steven, and Roxanne van Delft. 2015. "Publishing in Public Administration: Issues with Defining, Comparing, and Ranking the Output of Universities." *International Public Management Journal* 18 (1): 87–107. https://doi.org/10.1080/10967494.2014.972482.

Walker, Richard M, Gene A Brewer, M. Jin Lee, Nicolai Petrovsky, and Arjen van Witteloostuijn. 2019. "Best Practice Recommendations for Replicating Experiments in Public Administration." *Journal of Public Administration Research & Theory* 29 (4): 609–626. https://doi.org/10.1093/jopart/muy047.

Walker, Richard M., Oliver James, and Gene A. Brewer. 2017. "Replication, Experiments and Knowledge in Public Management Research." *Public Management Review* 19 (9): 1221–1234. https://doi.org/10.1080/14719037.2017.1282003.

Walker, Richard M., Dannii Yuen-Lan Yeung, M. Jin Lee, and Ivan P. Lee. 2020. "Assessing Information-Based Policy Tools: An Eye-Tracking Laboratory Experiment on Public Information Posters." *Journal of Comparative Policy Analysis: Research and Practice* 22 (6): 558–578. https://doi.org/10.1080/13876988.2020.1753035.

Webeck, Sean, and Sean Nicholson-Crotty. 2020. "How Historical and Social Comparisons Influence Interpretations of Performance Information." *International Public Management Journal* 23 (6): 798–821. https://doi.org/10.1080/10967494.2018.1550129.

Willems, Jurgen, Peter Rasmussen Damgaard, and Wouter Van Dooren. 2019. "What if Kahneman is Wrong? An Evaluation of the Use of Cognitive Theory in Behavioral Public Administration." *EGPA Conference 2019*, Belfast.

Willems, Jurgen, Carolin J. Waldner, and John C. Ronquillo. 2019. "Reputation Star Society: Are Star Ratings Consulted as Substitute or Complementary Information?" *Decision Support Systems* 124 (September): 113080. https://doi.org/10.1016/j.dss.2019.113080.

Williams, Ross, and Gaétan de Rassenfosse. 2016. "Pitfalls in Aggregating Performance Measures in Higher Education." *Studies in Higher Education* 41 (1): 51–62. https://doi.org/10.1080/03075079.2014.914912.