THOSE WHO UNDERSTAND IT WILL NOT BE PERSUADED: A PERFORMANCE INFORMATION PARADOX

RUNNING HEAD: A PERFORMANCE INFORMATION PARADOX

ABSTRACT

Performance information has been suggested as a means to inform citizens about—and shape their reactions to—public sector performance. However, individuals' ability to process information varies considerably. This implies that the same item of performance information may be understood differently by different citizens. Drawing on cognitive psychology research, we argue that cognitive differences affect citizens' ability to interpret performance information and moderate the extent to which performance information affects citizen satisfaction with public services. The argument is tested in a large-scale survey experiment. Our findings provide evidence of a performance information paradox: Those who are better able to interpret performance information do not adjust their satisfaction levels in response to the information. Conversely, those with lesser abilities to interpret and understand performance information are more responsive to it, but their responses are often based on invalid interpretations.

INTRODUCTION

Increasing citizens' knowledge by means of government information is often assumed to strengthen democratic control and to increase various aspects of citizen participation (Lupia 1994; Lupia and McCubbins 1998; James 2011). Political science and public administration studies have examined the extent to which factual information affects, for instance, citizen vote choice (James 2011), the trustworthiness of public organizations (Grimmelikhuijsen and

Meijer 2014), and the extent to which citizens engage in co-production (Thomsen and Jakobsen 2015).

Performance information is one specific kind of government information that is used massively at different levels of government (Radin 2006; James 2011; Van Dooren 2011). Performance information targeted at citizens in the form of benchmarking, ratings, rankings, and quality reports is typically published in order to increase transparency, assist citizens in holding governments and public organizations accountable for their performance, and to facilitate exit and voice mechanisms in the public sector (Le Grand 2010; Van de Walle and Roberts 2011: 221f; James and Moseley 2014). The underlying assumption is that an informed citizenry is better able to hold governments accountable for their performance and to react to poor performance either by expressing their opinion or finding a better provider. Performance information can thus be seen as a potentially important ingredient in assuring that NPM-inspired bottom-up models of public service delivery such as voice and choice will be effective.

However, research has shown that certain cognitive biases affect how citizens interpret information. Interpretations of performance information have been shown to be context-dependent in the sense that similar performance by public and private organizations are assessed differently (Hvidman and Andersen 2016; Marvel 2015), to be prone to an "odd-pricing" effect whereby one tenth of a point changes in ratings might sometimes lead to considerable changes in how people assess performance (Olsen 2013), to be asymmetrical in the sense that low performance receives stronger emphasis than positive performance (James and John 2007; Olsen 2015), and to be contingent upon citizens' prior beliefs (Baekgaard and Serritzlew 2016; Baekgaard et al. 2017). While the question of whether citizens generally respond in an unbiased manner to performance information thus seems to be well covered in

the literature, much less is known about whether different citizens respond differently to performance information.

The broader – largely experimental – literature on how people respond to factual information suggests that politically sophisticated individuals have a more nuanced understanding of information and hence that cognitive ability is an important determinant of how individuals engage with information (Mondak et al. 2007; Tilley and Hobolt 2011; Singh and Roy 2014). Moreover, research has shown that the ability to process numerical information varies considerably (Peters et al. 2006; Reyna et al. 2009). This implies that different citizens may understand the same item of information very differently. Since quantified performance information is a type of government information that puts particular demands on recipients in terms of processing and comprehending the information, the cognitive ability to process numerical information is likely to be of particular relevance to this kind of information.

Hence, we argue that cognitive differences moderate the extent to which performance information affects perceptions of public service performance. Specifically, we argue that performance information is less likely to have an impact among those with the strongest ability to understand and interpret numerical information, as such citizens have a more nuanced understanding of the information and assign less relevance to it than citizens with lesser abilities. We thus hypothesize a performance information paradox: Those who are better able to interpret performance information are less likely to respond to the information, and performance information therefore makes little difference for this group. Conversely, those who are less able to interpret and understand performance information are more likely to be responsive to the information, but their responses are often based on invalid interpretations of the information.

This idea is tested in a survey and a large-scale experiment with responses from a representative sample of Danish citizens. In the survey-based test, citizens were presented with numerical information on the performance of a public and a private organization. They were then asked which of the organizations was performing better. The information had one correct interpretation. As a measure of the ability to process numerical performance information, we use the so-called numeracy scale (Weller et al. 2013), which is the quantitative counterpart of literacy. The test confirms that the generally positive effect of numeracy on understanding identified in the literature (Peters et al. 2006; Reyna et al. 2009) also applies to the interpretation of performance information. In the experiment, citizens were randomly assigned to control and treatment groups. Those in the treatment group received an information cue about the average performance of public schools in their local government compared to the performance of schools in other local governments. They were then asked to indicate their satisfaction with the public school services of their local government. The experiment shows that the effect of receiving performance information is stronger for individuals with a weaker ability to understand and interpret numerical information, as measured by the numeracy scale.

The paper proceeds with a discussion of how and why citizens' ability to interpret performance information correctly is expected to vary. We introduce a measure of ability to process numerical information (numeracy) and discuss how it relates to citizens' interpretations of and attitudinal responses to performance information. We then present the design and data and discuss how to address the research question. Next, we present our empirical findings. We conclude by discussing the implications of our findings for research and practice.

COGNITIVE CAPABILITIES AND THE ABILITY TO PROCESS PERFORMANCE INFORMATION

Performance information is typically presented to citizens in the form of benchmarks, ratings, or rankings (Van de Walle and Roberts 2011), i.e., quantitative, numerical data on how an organization or service is performing compared to the past or compared to other organizations or services. Presenting performance information in this manner ideally allows citizens and other stakeholders, such as politicians and civil servants, to make informed decisions based on updated knowledge regarding performance (Moynihan 2008: 6; Van de Walle and Roberts 2011: 222). This requires that the recipients possess the necessary cognitive skills to understand and use the information (Reyna et al. 2009). Such skills, however, cannot be taken for granted. For instance, a large share of the US adult population has limited quantitative skills and cannot solve basic computational problems, for instance converting ratios to percentages (Kirsch et al. 2002).

Research on how people respond to factual information indicates that people assign varying relevance to different pieces of information when making judgments depending on their cognitive abilities (Mondak et al. 2007). More politically sophisticated individuals have been shown to have a more nuanced understanding of information and hence to be more likely to respond to such information (Tilley and Hobolt 2011; Singh and Roy 2014). Likewise, research has shown that the ability to process and use numerical information varies considerably and that numerical skills are important in terms of how individuals respond to information. The literature covers a wide range of responses to information moderated by numeracy, including risk perception, emotional response, and decision-making performance (Reyna et al. 2009). From a performance information perspective, the focus should be on whether recipients are able to interpret the information and whether they respond by changing attitudes or behavior (Baekgaard 2015).

In the following, we use insights from cognitive psychology to argue that the effect of performance information may be quite paradoxical: Recipients who are able to process numerical information (measured by numeracy) are more likely to understand performance information correctly, but recipients who have problems processing numerical information are more likely to be persuaded by it.

First, the effect of numeracy on the understanding of performance information: Cognitive psychology research shows how and under what circumstances numeracy is related to the comprehension of new information. In four experimental studies of the interplay of numeracy and affective considerations in shaping reactions to numerical information, Peters et al. (2006) find that highly numerate individuals tend to focus more on the details of numbers and are more likely than less numerate individuals to use appropriate numerical principles when interpreting numerical information. Peters et al. (2007) focus on how different information presentation methods affect how citizens comprehend health performance data. They find that experiment participants with higher numeracy are considerably better at comprehending the information regardless of its complexity. Kahan et al. (2017) show that numeracy has a strong positive impact on the ability to comprehend uncontroversial information about the performance of a skin cream, and a marginally positive impact on the comprehension of information of similar numerical strength about a gun ban. In conclusion, numeracy has been shown to have a positive impact on the comprehension of numerical information. Although the general effect is well known, we test this to confirm that the basic logic also applies to performance information and to validate our measure of numeracy. We find that individuals with better numerical skills are on average better able to process and comprehend quantitative information.

The second question is whether individuals are persuaded by the information, or, more specifically, whether the effect of information on satisfaction is related to numeracy. A high

ability to process numerical information implies that individuals are better able to understand the information and to use quantitative data more effectively to arbitrate among competing claims. This has been pertinent in the health care utilization literature. For example, patients are often confronted with quantitative information about their ailment. Highly capable patients are better able to understand this information, which renders them better able to react to the information, and, ultimately, to receive better care. Consistent with this, Abdel-Kader et al. (2010) find that numeracy is important to the care for kidney patients. Patients with higher numeracy are more likely to receive a kidney transplant, independent of sex, race, and age. The study shows that numeracy does have real effects on outcomes that are important to individuals. Furthermore, highly numerate individuals may have greater faith in numerical information than less numerate individuals. Gurmankin, Baron, and Armstrong (2004) conducted an experiment in which they compared the trust and confidence of the participants in the information provided in three scenarios in which they were presented with either verbal information or one or two pieces of numerical information. They found that the highly numerate participants trusted the information in the numeric formats more, whereas those with the lowest numeracy scores trusted the information in the verbal format more. When individuals trust the information more, they are also more likely to be comfortable using it. However, understanding, trusting, and being able to act upon new information does not imply that new performance information affects opinion.

There are two reasons to expect that high numeracy is likely to be associated with a reduced effect of information on opinion. First, highly numerate individuals are better able to assess and aggregate information and are therefore likely to have more knowledge about an issue before being exposed to new performance information. This implies that we should expect new information to matter less for this group. Second, more numerate individuals are likely to treat new information as updates of knowledge. The probabilogical model of belief

change (Eagly and Chaiken 1993: 221-231) offers an account of how changes in individuals' beliefs about the premises of some conclusion affect their belief about the conclusion. This is the exact same question we are interested in here: Performance information can potentially change beliefs about some of the premises of an individual's opinion about, for instance, the quality of a government service. According to the probabilogical model, individuals' opinions will change if their beliefs about one of the premises of the opinion changes. However, they will not simply assume that a premise is either true or false but tacitly assign a probability that it is true (Jost and Hardin 2011: 28). Using an example that will reappear in the experiments presented below, an individual's evaluation of the performance of a school is expected to change if one of the premises of this evaluation (in this example information about grades compared to the national average) changes. However, new information does not automatically produce a shift in beliefs about a premise. Building upon Bayesian statistics, Fischhoff and Beyth-Marom (1983: 240) argue that new information should be thought of as updates of prior knowledge. Optimal behavior implies that the individual assigns appropriate weights to prior knowledge and to new pieces of information. This means that a new item of information (e.g., on the grades given in a school) should not automatically produce a shift in beliefs about attainment levels, let alone in the overall opinion of the quality of the school. Rather, individuals with reliable prior knowledge concerning the specific school should be less likely to be affected by new information than individuals with less reliable prior knowledge.

However, individuals can fail to assign appropriate weight to prior knowledge and new information. For instance, the base-rate fallacy implies that too much weight is assigned to new information, while prior knowledge is ignored. According to Kahneman and Tversky, this tendency is common in human intuition: "The failure to appreciate the relevance of prior probability in the presence of specific evidence is perhaps one of the most significant departures of intuition from the normative theory of prediction" (1973: 243). Numeracy is

important in terms of how new information is understood. Hart (2013) finds that less numerate individuals are relatively more susceptible to numeric descriptors. When presented with numeric information regarding the effects of climate change, less numerate individuals became significantly more concerned for victims and more willing to donate money. Conversely, highly numerate individuals were not significantly affected by the information. Moreover, Peters et al. (2006) show how highly numerate individuals are generally more detail-oriented and therefore better able to see through the deficiencies of information, so we might expect them to be more skeptical regarding performance information, which is often relatively simple and only covers one or a few performance indicators, whereas costs and quality are often hardly mentioned if at all. If those with a better understanding of quantitative information are (1) more knowledgeable before new information is presented and (2) less prone to the base-rate fallacy, new information will have relatively *less* effect on the opinions of numerate individuals. In other words, we expect proficiency to process numerical information to be associated with better understanding of performance information but with the information having less impact on opinions. We call this the persuasion hypothesis.

Performance information has less impact on opinion for individuals with higher numeracy

In other words, we expect to find a performance information paradox: Performance information is best understood by those with higher numerical skills but has a greater impact on those with lesser skills; that is, those who are more likely to misunderstand the information.

DATA AND DESIGN

To test the persuasion hypothesis, we investigate how performance information exposure affects opinion for people with different degrees of numeracy. Moreover, to ascertain that numeracy is indeed a valid measure of the ability to comprehend performance information, we have to conduct a separate test of whether the respondents' ability to process numerical information (measured as numeracy) affects their comprehension of performance information.

In order to eliminate potential simultaneity bias, control for confounders, and control which information is provided to whom, we follow the standard in the opinion formation literature (see, e.g., Chong and Druckman 2007: 108) and use survey experiments (see also Baekgaard and Serritzlew 2016 for a similar application). We study the two sides of the paradox in two separate survey experiments. Both are based on a 2014 survey of 1,784 Danish citizens recruited from an Internet panel administered by the Danish survey company Userneeds. While the panel is representative of the Danish working-age population at large on background characteristics such as gender, age, level of education, and geographical affiliation, survey respondents are arguably likely to differ from the population at large with regard to for instance political interest.[1] Important for our purposes, however, is that while we cannot be certain that the findings generalize to the population at large on unobservable characteristics, our ideas are tested on a very diverse sample, thus increasing our confidence in the external validity. Another question is to what extent results based on Danish citizens can be generalized to other countries. Denmark is a developed welfare state with a high level of spending on education, a relatively well-educated population, and strong support of the public sector. This may affect the variation in numeracy and perhaps lead to generally more positive views on public institutions (which is related to the dependent variable in the two studies presented below). It is possible that the effects of numeracy will be larger in countries with a higher variation in numeracy and with more diverse views on public institutions.

Findings from the survey experiments have previously been reported in a study of the effect of the interpretation of performance information (Baekgaard and Serritzlew 2016) and a study of whether the effect of performance information on the motivation to co-produce is mediated by satisfaction (Blom-Hansen and Baekgaard 2015). Here, data is used for a different purpose than in the articles mentioned above, as our focus is on numeracy and how it affects the outcome variables of the two experiments. The experiments are described in detail below.

The survey method also allows us to use a standard measure of numeracy. Numeracy can be defined as the ability to understand and use numbers (Reyna et al. 2009). The concept is complex and encompasses dimensions requiring various skills. At the most fundamental level, "skills associated with numeracy include the ability to perform simple arithmetic operations and compare numerical magnitudes. At a higher level, numeracy encompasses basic logic and quantitative reasoning skills, knowing when and how to perform multistep operations, and an understanding of ratio concepts, notably fractions, proportions, percentages, and probabilities" (Reyna et al. 2009: 947). We measure numeracy using a modified version of an eight-item psychometric measure: The Abbreviated Numeracy Scale (Weller et al. 2013).[2] This measure involves a test of the respondents' numerical abilities, and numeracy is hence measured as the number of correct answers to the questions in Table 1.

Table 1 here

Roughly 6% of the respondents gave no correct answers, and around 9% answered all eight questions correctly.[3] With a mean value of 4.7 and a standard deviation of 2.3, numeracy varies reasonably among the respondents. Figure 1 depicts the distribution of the numeracy variable. Unsurprisingly, as shown in Table OA1 in the online appendix, highly

numerate individuals differ from the least numerate on background characteristics. Highly numerate individuals tend to be male, better educated and have a higher income. Importantly, the least numerate do not show less engagement in responding to the numeracy question. In fact, the least numerate tend to spend slightly more time on completing the questionnaire than the most numerate.

Table 2 shows the questionnaire design. Numeracy was measured late in the survey after the respondents were introduced to the experimental treatments and after the dependent variable used to test the persuasion hypothesis was measured. We decided on this design because we were concerned that early exposure to the numeracy questions could potentially prime people to be more careful in how they responded to our experimental treatments. From a more practical perspective, we were also concerned that early exposure to numeracy would create a large drop-out from the survey. The questionnaire design does, however, raise the question whether numeracy has somehow been affected by the experimental treatments. At the outset, numeracy should be considered a trait that is difficult to manipulate. Using dummy regression in which each experimental arm is assigned a dummy category, we indeed find no significant impact of the experimental treatments on numeracy (see table OA4 and OA5 in the online appendix).

Table 2 here

The Effect of Numeracy on Comprehension

Baekgaard and Serritzlew (2016) show in a survey experiment that interpretation of performance information is subject to motivated reasoning: The interpretation of new performance information is systematically biased by prior ideological beliefs. In the experiment, respondents were presented with performance information with only one correct

interpretation. However, the information did require some processing: Respondents were shown the number of successful operations at two hospitals and asked to assess which of the two was performing better. Correct interpretation of the performance information requires a rough understanding of the relative incidence of complications from the absolute numbers. The experiment showed that prior beliefs affect how the unambiguous information is interpreted.

We use data from this experiment (the dependent variable) to verify that numeracy has a positive effect on comprehension of performance information. This serves as a check of the basic logic of the study and as a validity check of the measure of numeracy. Since the performance information provided is unambiguous, respondents' answers about the relative performance are either right or wrong. We expect that highly numerate respondents are more likely to give correct answers. The respondents in the treatment groups were presented with performance information (complications in hip operations) for either a public or a private hospital. The experiment is presented in Appendix A. We investigate the effect of numeracy in a logistic regression with a dummy variable indicating whether or not the respondent was able to understand the performance information (i.e., was able to answer the performance information question correctly) as the dependent variable and with numeracy as the independent variable. Since we do not randomize numeracy (this is a trait of each individual that cannot be directly manipulated), control variables must be included in order to account for potential spuriousness. We control for the following individual-level characteristics: Gender, age, level of education, income, and party ideology. Key descriptive statistics for the variables used in the analyses are presented in Table 3,[4] and the distribution of key variables is presented in Figure 1.

Table 3 here

Figure 1 here

Table 4 shows logistic regression analyses of the effect of numeracy on the probability of answering the question regarding the performance of the two hospitals correctly. Model 1 shows the results of a bivariate analysis of the relationship between numeracy and the ability to interpret the information correctly. As expected, the two variables are strongly positively correlated. The coefficient corresponds to a 30 percentage point difference in the predicted probability of responding correctly between those with the lowest (49.9% responded correctly) and highest numeracy (79.8% responded correctly).

Table 4 here

Model 2 includes the individual level controls and control for the extent to which the information is politicized. As expected, the numeracy coefficient continues to be significant and of almost the same size as in the analysis without controls.[5] Respondents with higher numeracy are better able to interpret performance information correctly. We infer that the well-known effect of numeracy also applies to performance information, and we take the results as a validation of numeracy as a useful measure of numerical competence.

Persuasiveness of Performance Information

We now proceed to test the persuasion hypothesis, i.e. whether numeracy moderates the effect of performance information on how respondents assess a public organization. Blom-Hansen and Baekgaard (2015) investigate how performance information about Danish public schools in the respondent's municipality affects their satisfaction with the schools. The Danish public sector is highly decentralized, and elected local politicians have autonomy to make decisions

of relevance to the schools within their jurisdiction, about school structure, governing of public schools, and within the limits of the law, the number of weekly lessons in each subject. Respondents were initially divided into two groups depending on the municipal school performance. They were later randomly assigned to a neutral treatment and a treatment with real information about the performance of public schools in the municipality. This allows for a comparison of the impact of information among respondents from municipalities performing at the same level. A similar design is used in Baekgaard (2015), Nielsen and Baekgaard (2015), and George et al. (2017). Table 5 shows the treatments.

Table 5 here

In treatment 1, the respondents received no performance information; in treatment 2, they received information on the actual performance (in the top or bottom-third of Danish municipalities) of the public schools within their own municipality. Blom-Hansen and Baekgaard (2015) show how actual public school performance is positively correlated with higher satisfaction—but only among those who received performance information. From this, they conclude that performance information does affect satisfaction.

We are interested in a related question. We know that performance information affects satisfaction, but according to the persuasion hypothesis, performance information should have a stronger effect on respondents with low numeracy. We use the same experimental setup, reuse data on the dependent variable (satisfaction), add numeracy, and analyze the effect of the performance information treatment and its interaction with numeracy (measured as described above) on satisfaction. Satisfaction was measured using the question: "The quality of the public schools in my municipality is satisfactory." Response options were complete

agreement, agreement, neither agreement nor disagreement, partial disagreement, or complete disagreement.

Since numeracy is not manipulated experimentally, we include control variables to account for alternative explanations. Education (whether the respondent has a university degree) is included to capture the general knowledge of the individual acquired from education, and user status (whether the respondent has at least one child enrolled in a public school) to capture their specific knowledge generated by personal experience with the service.[6] We also include three socio-economic variables that could potentially be related to general political attitudes. Finally, since the information provided in the treatment group was based on actual performance, municipality-fixed effects are included to account for cross-municipal differences.

Table 6 shows the results of ordered logistic regression analyses, the five-point satisfaction scale serving as dependent variable.[7] The expected moderated effect is tested by an interaction term between information and numeracy. Model 1 and Model 2 test the effect for citizens living in high-performing and low-performing municipalities, respectively.[8]

Table 6

As we should expect, Model 1 shows that performance information has a positive impact on the satisfaction among citizens residing in high-performance municipalities. Moreover, the negative and statistically significant interaction term between information and numeracy shows that information has a stronger impact the lower the numeracy of the respondents. Consistent with the persuasion hypotheses, this finding supports the idea that numeracy curbs the effect of performance information on attitudes. Model 2 shows a similar, although much weaker, pattern for the respondents living in low-performance municipalities. Receiving the

performance information reduces satisfaction, but the effect is smaller than in model 1. The effect seems to be stronger for low-numeracy respondents (the point estimate of the effect of performance information for respondents with minimum numeracy is – 0.730; for respondents with maximum numeracy it is –0.730 + 8 × 0.050 = –0.330). However, the interaction effect is not statistically significant in this model.

The fact that numeracy and its interaction with the treatment variable are significant in model 1 is consistent with the persuasion hypothesis. To get an impression of the levels where numeracy information has an impact on satisfaction, we visualize the findings from Table 6 in Figure 2.

Figure 2 here

Figure 2 shows the marginal effects of information for different levels of numeracy. The figure is a simplified version of the analysis in Table 6 in the sense that we look at the marginal effect of information on being 'very satisfied'. Panel A shows the relationship for the high performance treatment; panel B for the low performance treatment. Panel A shows that the impact of information varies substantially across individuals with different levels of numeracy. While information indicating high performance increases the predicted probability of being very satisfied significantly with about 10 percentage points for the least numerate, the impact of information is close to zero and clearly insignificant for the most numerate. In panel B, the slightly positive slope indicates that the impact of information is almost similar for all individuals regardless of how numerate they are. This is consistent with the findings in table 6, model 2. However, the panel also provides some evidence that numeracy actually matters since the impact of information is negative and significant for the least numerate and insignificant for the most numerate individuals. Overall, the findings are consistent with the

persuasion hypothesis: The opinions of highly numerate respondents are less affected by exposure to performance information.

DISCUSSION AND CONCLUSION

Use of performance information has become fashionable in the public sector. However, politicians and citizens do not automatically process performance information perfectly. The ability to process numerical information (measured as numeracy) varies substantially (Peters et al. 2006; Gigerenzer et al. 2007; Reyna et al. 2009).We argue that cognitive ability is decisive for whether performance information is understood and changes opinions.

We expect performance information to be less likely to have an impact on opinion among the highly numerate, i.e. those who are best able to understand the information. This leads to a performance information paradox: Those who understand performance information will not be persuaded. In other words, those who are better able to interpret performance information are unlikely to respond to the information, and performance information therefore makes little difference for this group. Conversely, those who are less able to interpret and understand performance information are more likely to be responsive to the information, but their responses are often based on invalid interpretations of the information.

We investigate this by means of a survey-based test and a survey experiment with responses from a representative sample of Danish citizens. We find strong support for the well-known effect of numeracy on the ability to understand information. Highly numerate respondents are better able to understand the performance information presented to them. The results of the survey experiment are consistent with the persuasion hypothesis: the opinions of highly numerate respondents are less affected by exposure to performance information. This effect can be due to two mechanisms. First, highly numerate individuals are, according to the

probabilogical model of belief change, better able to appropriately discount new information. Second, highly numerate individuals are likely to be more knowledgeable about an issue before being exposed to new performance information, and they may also generally be more skeptical of the validity of standardized tests. Moreover, they may have high expectations to the school they have chosen for their children and therefore less impressed by the test results. Which of these mechanisms is strongest remains an open question. The present survey instrument does not allow us to distinguish between these two different mechanisms. In the theoretical section we argue, for instance, that numeracy is associated with being more detail oriented and with more reflective thinking and therefore with a better ability to identify deficiencies of numerical information.

Further research could in similar designs test this mechanism simply by adding questions on to what extent the respondents consider the information provided a useful and complete indicator of performance. According to the first mechanism, highly numerate individuals should indicate that the information is less useful and complete. Questions on respondents' view of standardized tests and their expectation of the quality of their school could clarify whether highly numerate differ from less numerate individuals. If the second mechanism is relevant, the effect of numeracy should decrease when we control for these measures.

It is also relevant to consider the extent to which the results can be generalized to other countries and to other types of performance information than what has been presented in this study. The high education levels in Denmark may imply that the variation in numeracy is relatively small, and the generally strong support of the public sector may affect perceptions of the performance of public and private organizations. It is therefore possible that the substantive effect of numeracy will be larger in countries with higher variation in these variables. Regarding the type of information, the experiments use quantitative and stylized data. Future research could investigate the performance information paradox with other types

of performance information. This is particularly relevant because numerical skills have been shown to correlate with more reflective thinking (Campitelli and Labollita 2010) and thus with performance in non-numerical tasks. Since cognitive reflection is partly contained in the numeracy measure that we use, we would indeed expect similar effects for non-numerical performance information.

Since the results are based on an experiment, the external validity of the results is limited. Moving beyond survey experiments could show whether performance information presented through usual means is susceptible to the paradox. Along the same line, we would expect a stronger impact of numeracy, the more challenging and complex the numerical information. Future research might therefore benefit from manipulating the complexity of numerical information in order to investigate whether this is indeed the case. Also, the information provided in this study has a one-shot character. However, highly numerate individuals may need more information to be persuaded and may therefore be more responsive to performance information after repeated exposure. Conversely, repeated exposure might lead to a more nuanced understanding and hence less strong responses among less numerate individuals. Future research might therefore benefit from studying how individuals react to repeated exposure to performance information. Finally, we study the two sides of the performance information paradox in two separate studies. The design can be strengthened by integrating the two tests into a single experiment.

Our research contributes in two ways to the extant body of research on the effects of performance information targeted at citizens. First, recently published articles provide evidence that various cognitive biases affect how citizens process and understand information (e.g. Marvel 2015; Olsen 2015; Hvidman and Andersen 2016). By identifying numeracy as an explanation of how individuals respond to information, this research not only lends further support to the idea that cognitive biases play a central role in how individuals process

information but also suggests an important individual-level trait that might help understanding why some citizens are prone to other biases such as the so-called "odd-pricing effect" (Olsen 2013).

Second, and more importantly, the findings illustrate that citizens respond fundamentally differently to performance information. The performance information paradox therefore questions the universal usefulness of performance information. One should not assume that everyone understands performance information. On the contrary, fully understanding it requires skills.

The performance information paradox has implications for the use of performance information in the public sector in three important ways. First, it points to a problem of equity. Second, it questions whether mere publication of performance information can be an effective driver of change, and this suggests managerial attention is warranted if performance information is supposed to matter. Third, it suggests that the way performance information is presented is likely to matter.

The question of equity is important because information about public performance has potential private benefits to citizens. Performance information about services can have direct consequences for choice. Citizens who are able to understand and act upon performance information will be able to select government services of higher quality, i.e. send their children to better public schools or receive treatment at better public hospitals. The ideal that citizens should be able to benefit equally from government services and initiatives is widely accepted (Boyne et al., 2003), and the performance information paradox implies that we cannot take for granted that all groups of citizens are equally able to benefit from it.

The question of performance information as a driver of change is relevant because the performance information paradox implies that performance information does not automatically affect opinion, particularly among those who understand the information best.

If the information has limited effects on opinion, it is also likely to have limited effects on behavior such as citizens' use of voice and choice (Hirschman, 1970). Managers will therefore have to consider how performance information can be supplemented by strategies aimed at ensuring appropriate reactions to poor performance.

The fact that groups of citizens have difficulties understanding performance information also suggests that the way information is presented is important. Performance information should be provided in a simple fashion that allows citizens with different backgrounds to benefit from it. Before adopting ambitious performance information systems designed to inform the public about the quality of the public sector, managers should consider whether the information is likely to be understood by all; and if not, by whom. It is also important to keep in mind that performance information is potentially harmful in the sense that it will be routinely misunderstood and that it is quite possible that it will affect the opinions of the very citizens who are most likely not to fully understand the information.

The research also contributes to the broader literature on how people respond to factual political information in which it has several times been shown that more politically sophisticated individuals – typically measured by their level of general political knowledge or general political interest – are more likely to change their opinions in response to such information (Tilley and Hobolt 2011: 5-6; see also Grimmelikhuijsen and Meijer 2014 for a notable exception). It is remarkable that our findings are in outright opposition to the general trend in this literature in the sense that political sophistication – measured by numeracy – curbs the impact of performance information on satisfaction. The discrepancy highlights that research needs to get a better grasp of why political knowledge matters to the effect of information in some cases but not in others and of the extent to which various dimensions of sophistication (such as knowledge and numeracy) are related or separate phenomena.

NOTES

[1] Userneeds does not allow self-registration to their panel. They approach people with different background characteristics to ensure that their panel is as representative on socio-demographic characteristics as possible.

[2] Six of our items are slightly modified versions of six of the items in the eight-item index suggested by Weller et al. (2013), the only difference being that prices are written in Danish currency instead of US dollars. We replaced Q8b in Weller et al. (2013) with Q10, because Q8b cannot stand alone but requires that a question not included in the eight-item measure has been asked previously. Q8b and Q10 both measure the ability to calculate probabilities; moreover, Q12, which requires a demanding graphical presentation, was replaced with Q16.

[3] One potential concern is that some less numerate respondents did not provide valid responses due to low engagement. We checked the potential impact of low engagement on our results by re-coding the numeracy variable as missing for all respondents who either did not fill in numerical responses for at least four or more of the numeracy items or gave the exact same numerical response to all eight numeracy items. Based on these criteria, 89 responses are coded as missing. As can be seen from the robustness checks in the online appendix in Table OA6 (model 1) and Tables OA7 and OA8 (in both cases model 5), our findings become even stronger when the recoded numeracy variable is used.

[4] Full descriptives for all models can be found in Table OA2 and OA3 in the online appendix.

[5] As described in Appendix A, the survey experiment had politicized and neutral treatments. It turns out that numeracy has a stronger effect in the neutral variant, but the effect is strong in both groups (in the politicized treatments the predicted probabilities of giving a correct answer vary between 53.9% among the least numerate and 75.6% among the most numerate respondents (see also model 2 in Table OA6 in the online appendix).

[6] The findings are robust to other operationalizations of user status.

[7] We conducted a series of additional analyses to check the robustness of our findings. In all the checks the findings are substantially similar to those in Table 6. Among the checks we conducted are

regressions in which additional interaction terms between performance information and education and performance information and user status, respectively, are included. The robustness checks are reported in the online appendix in Table OA7 and OA8.

[8] A third group received information that their municipalities performed about average. Since we have no *a priori* expectations about the impact of information on satisfaction for this particular group, we refrain from presenting the findings for these individuals.

REFERENCES

Abdel-Kader, K., M. A. Dew, M. Bhatnagar, C. Argyropoulos, I. Karpov, G. Switzer, and M. L. Unruh. 2010. "Numeracy Skills in CKD: Correlates and Outcomes." *Clinical Journal of the American Society of Nephrology* 5: 1566-1573.

Baekgaard, M. 2015. "Performance Information and Citizen Service Attitudes: Do Cost Information and Service Use Affect the Relationship?" *International Public Management Journal* 18: 228-245.

Baekgaard, M. and S. Serritzlew. 2016. "Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension?" *Public Administration Review* 76: 73-82.

Baekgaard, M., J. Christensen, C. M. Dahlmann, A. Mathiasen, and N. B.G. Petersen. 2017. "The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Politicians." *British Journal of Political Science*.

Blom-Hansen, J. and M. Baekgaard. 2015. "Kan borgerne motiveres til at samproducere ved hjælp af information om performance?" *Politica* 47: 144-163.

Boyne, G., C. Farrell, and J. Law. 2003. *Evaluating Public Management Reforms: Principles and Practice*. Open University Press.

Campitelli, G. and M. Labollita. 2010. “Correlations of Cognitive Reflection with Judgments and Choices.” *Judgment and Decision Making* 3: 182-191.

Chong, D. and J. N. Druckman. 2007. “Framing Theory”. *Annual Review of Political Science* 10: 103-126.

Eagly, A. H. and S. Chaiken. 1993. *The Psychology of Attitudes*. Fort Worth: Harcout Brace College Publishers.

George, B., S. Desmidt, P. A. Nielsen, and M. Baekgaard. 2017. “Rational Planning and Politicians’ Preferences for Spending and Reform: Replication and Extension of a Survey Experiment.” *Public Management Review* 19(9): 1251-1271.

Gurmankin, A. D., J. Baron, and K. Armstrong. 2004. “The Effect of Numerical Statements of Risk on Trust and Comfort with Hypothetical Physician Risk Communication.” *Medical Decision Making* 24: 265-271.

Fischhoff, B. and R. Beyth-Marom. 1983. “Hypothesis Evaluation from a Bayesian Perspective.” *Psychological Review* 90: 239-260.

Gigerenzer, G., W. Gaissmaier, E. Kurz-Milcke, L. M. Schwartz, and S. Woloshin. 2007. “Helping Doctors and Patients Make Sense of Health Statistics.” *Psychological Science in the Public Interest* 8: 53-96.

Grimmelikhuijsen, S. G. and A. J. Meijer. 2014. “The Effects of Transparency on the Perceived Trustworthiness of a Government Organization: Evidence from an Online Experiment.” *Journal of Public Administration Theory and Research* 24: 137-157.

Hart, P. S. 2013. “The Role of Numeracy in Moderating the Influence of Statistics in Climate Change Messages.” *Public Understanding of Science* 22: 785-798.

Hirschman, Albert O. 1970. *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Boston: Harvard University Press.

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

James, O. 2011. “Performance Information and Democracy: Information Effects on Citizens in Field and Laboratory Experiments.” *Journal of Public Administration Research and Theory* 21: 399-418.

James, O. and P. John. 2007. “Public Management at the Ballot Box: Performance Information and Electoral Support for Incumbent English Local Governments.” *Journal of Public Administration Research and Theory* 17: 567-580.

James, O. and A. Moseley. 2014. “Does Performance Information about Public Services Affect Citizens’ Perceptions, Satisfaction, and Voice Behaviour? Field Experiments with Absolute and Relative Performance Information.” *Public Administration* 92: 493-511.

Jost, J. T. and C. D. Hardin. 2011. “On the Structure and Dynamics of Human Thought: The Legacy of William J. McGuire for Social and Political Psychology.” *Political Psychology* 32: 21-57.

Kahan, D. M., E. Peters, E. C. Dawson, and P. Slovic. 2017. “Motivated Numeracy and Enlightened Self-government.” *Behavioral Public Policy* 1: 54-86.

Kahneman, D. and A. Tversky. 1973. “On the Psychology of Prediction.” *Psychological Review* 80: 237-251.

Kirsch, I. S., A. Jungeblut, L. Jenkins, and A. Kolstad. 2002. “Adult Literacy in America: A First Look at the Findings of the National Adult Literacy Survey.” *NCES Report No. 1993-275. 3*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Le Grand, J. 2010. “Knights and Knaves Return: Public Service Motivation and the Delivery of Public Services.” *International Public Management Journal* 13: 56-71.

Lupia, A. 1994. “The Effect of Information on Voting Behavior and Electoral Outcomes: An Experimental Study of Direct Legislation.” *Public Choice* 78: 65-86.

Lupia, A. and M. D. McCubbins. 1998. *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge: Cambridge University Press.

Marvel, J. D. 2015. “Unconscious Bias in Citizens’ Evaluations of Public Sector Performance.” *Journal of Public administration Research and Theory*, E-publication ahead of print. doi: 10.1093/jopart/muu053.

Mondak, J. J., E. C. Carmines, R. Huckfeldt, D. G. Mitchell, and S. Schraufnagel. 2007. “Does Familiarity Breed Contempt? The Impact of Information on Mass Attitudes toward Congress.” *American Journal of Political Science* 51: 34-48.

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

Nielsen, P. A. and M. Baekgaard. 2015. Performance Information, Blame Avoidance, and Politicians’ Attitudes to Spending and Reform. *Journal of Public Administration Research and Theory* 25: 545-569.

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: 365-371.

Olsen, A. L. 2015. “Citizen (Dis)satisfaction: An Experimental Equivalence Framing Study.” *Public Administration Review* 75: 469-478.

Peters, E., N. Dieckmann, A. Dixon, J. H. Hibbard, and C. K. Mertz. 2007. “Less Is More in Presenting Quality Information to Consumers.” *Medical Care Research and Review* 64: 169-190.

Peters, E., D. Västfjäll, P. Slovic, C.K. Mertz, K. Mazzocco, and S. Dickert. 2006. “Numeracy and Decision Making.” *Psychological Science* 17: 407-413.

Radin, B. A. 2006. *Challenging the Performance Movement: Accountability, Complexity, and Democratic Values*. Washington, DC: Georgetown University Press.

Reyna, V. F., W. L. Nelson, P. K. Han, and N. F. Dieckmann. 2009. "How Numeracy Influences Risk Reduction and Medical Decision Making." *Psychological Bulletin* 135: 943-973.

Singh, S. P. and J. Roy. 2014. "Political Knowledge, the Decision Calculus, and Proximity Voting." *Electoral Studies* 34: 89-99.

Thomsen, M. K. and M. Jakobsen. 2015. "Influencing Citizen Coproduction by Sending Encouragement and Advice: A Field Experiment." *International Public Management Journal* 18: 286-303.

Tilley, J., and S. B. Hobolt. 2011. "Is the Government to Blame? An Experimental Test of How Partisanship Shapes Perceptions of Performance and Responsibility." *The Journal of Politics* 73: 1-15.

Van de Walle, S. and A. Roberts. 2011. "Publishing Performance Information: An Illusion of Control?" Pp. 215-230 in W. van Dooren and S. van de Walle, eds., *Performance Information in the Public Sector: How It Is Used*. Houndmills, Basingstoke: Palgrave Macmillan.

Van Dooren, W. 2011. "Nothing New Under the Sun? Change and Continuity in the Twentieth-century Performance Movements." Pp. 15-27 in W. van Dooren and S. van de Walle, eds., *Performance Information in the Public Sector: How It Is Used*. Houndmills, Basingstoke: Palgrave Macmillan.

Weller, J. A., N. F. Dieckmann, M. Tusler, C. K. Mertz, W. J. Burns, and E. Peters. 2013. "Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach." *Journal of Behavioral Decision Making* 26: 198-212.

Appendix A: Survey experiment testing the effect of numeracy on comprehension of performance information[1]

Table A1 shows the two versions of the performance information. Respondents were asked to evaluate the performance of the organization. In both treatments, they were given the following instructions:

> "We will now ask you to consider the following constructed example. Below is a table with information on the quality of hip operations at two hospitals—one public and one private. The patient groups at the two hospitals are very similar. The table shows how often hip operations were performed at the two hospitals with and without complications. Which of the hospitals performed best?"

Table A1 here

"Don't know" was not a response option in any of the cases. The respondents were thus forced to give their best estimate of which hospital was performing best. In Treatment 1, the private hospital performed better than the public one (complications in 16% compared to 25% of the cases). Treatment 2 produces the opposite result. For each respondent, we coded whether they provided the correct or incorrect answer. In both treatments, respondents were randomized into receiving either a politicized or a neutral identifier of the hospital. In the politicized version, the hospitals were referred to as "public hospital" or "private hospital," in the neutral version to "Hospital A" and "Hospital B". The use of different versions allows us to control effectively for whether numeracy has different effects depending on the extent to which the information is politicized.

[1] The appendix builds on Baekgaard and Serritzlew (2016). The experiment is inspired by Kahan et al. (2017).

Tables and figures

Table 1: Measuring numeracy

Q1. Imagine that we roll a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up as an even number?
Q2. In a lottery, the chances of winning DKK 100 are 1%. What is your best guess about how many people would win the 100 DKK prize if 1000 people each buy a single ticket?
Q3. In a lottery, the chance of winning a car is 1 in 1000. What percentage of tickets in the lottery win a car?
Q9. If the chance of getting a disease is 20 out of 100, this would be the same as having a _____% chance of getting the disease.
Q10. The chance of getting a viral infection is 10%. Out of 1000 people, approx. how many of them are expected to get infected?
Q15. A bat and a ball cost DKK 11 in total. The bat costs DKK 10 more than the ball. How much does the ball cost?
Q16. If it takes five machines five minutes to make five widgets, how long would it take 100 machines to make 100 widgets?
Q17. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

The question numbers in the battery refer to the question number in Weller et al. (2013). The battery was introduced with the following sentence: “We now ask you to give your best estimate of the correct answer to the following questions:”

Table 2: Questionnaire Design

Question no.	Question content
1-9	Questions unrelated to the content of this article (not used here)
Introductory text to question no. 10	School performance information treatment (the independent variable)
10	Satisfaction with public schools (the dependent variable)
11-24	Questions unrelated to the content of this article (not used here)
25	Experimental data used to measure comprehension of performance information focusing on hospital performance
26	Numeracy (the moderating variable)
27-33	Socio-demographic characteristics and party choice

Table 3: Descriptive statistics

Variable	Description	Mean	SD	Min	Max
Comprehension of information	Dependent variable in test of numeracy measure; Dummy variable; Correct responses coded 1	0.7	0.5	0	1
Citizen satisfaction	Dependent variable in the survey experiment; 5-point scale; Higher values correspond to higher satisfaction	3.2	1.1	1	5
Numeracy	Key explanatory variable in our experiments; Continuous variable	4.7	2.2	0	8
Gender	Dummy variable; Males coded 1	0.5	0.5	0	1
Age	Measured in years	43.8	14.2	18	67
Education	Dummy variable; University degree coded 1	0.6	0.5	0	1
Income	Dummy variable; Income above 400,000 coded 1	0.3	0.5	0	1
Ideology	Dummy variable; Voted for left-wing party at latest municipal election coded 1	0.5	0.5	0	1
User	Parents of one or more children enrolled in the public school in their home local government coded 1	0.2	0.4	0	1

Table 4: The impact of numeracy on the comprehension of information (logistic regression)

	Model 1	Model 2
Numeracy	0.197**	0.198**
	(0.022)	(0.028)
Politicized information		–0.073
		(0.115)
Gender		0.048
		(0.119)
Age		0.003
		(0.004)
Education		0.207+
		(0.124)
Income		–0.076
		(0.134)
Ideology		0.007
		(0.118)
Intercept	–0.201+	–0.410
	(0.109)	(0.258)
N	2,002	1,457
Nagelkerke's R^2	0.058	0.063
Begin-log likelihood	–1,272	–920
End-log likelihood	–1,229	–886
% correctly predicted	66	68

Entries are logistic regression coefficients; Robust standard errors in parentheses
$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

Table 5: Treatments in the survey experiment

Treatment 1: No performance information	Treatment 2: Performance information
There is often discussion about the quality of public services. For example, there has been much focus on how to measure the quality of education in the Danish public schools. To what extent do you agree or disagree with the following statements about the public schools?	There is often discussion about the quality of public services. For example, there has been much focus on how to measure the quality of education in the Danish public schools. According to a quality measurement from 2012, the GPA from the school-leaving examinations at the public schools in [insert name of respondent municipality] was in the [top/bottom] third among all municipalities when the social background of the individual pupil is accounted for. To what extent do you agree or disagree with the following statements about the public schools?

Table 6: The impact of performance information and numeracy on satisfaction

	Model 1: Information on high performance	Model 2: Information on low performance
Performance information	1.259*	–0.730*
	(0.551)	(0.348)
Numeracy	0.146+	0.009
	(0.084)	(0.059)
Performance information × numeracy	–0.200*	0.050
	(0.085)	(0.079)
Gender	–0.148	–0.154
	(0.214)	(0.206)
Age	–0.005	0.001
	(0.009)	(0.007)
Education	0.329	–0.020
	(0.232)	(0.157)
Income	0.124	–0.106
	(0.267)	(0.192)
Ideology	–0.041	0.235
	(0.252)	(0.226)
User	0.104	0.121
	(0.226)	(0.278)
N	425	336
Nagelkerke's R^2	0.168	0.122
Begin-log likelihood	–575	–497
End-log likelihood	–539	–477

Entries are ordered logistic regression coefficients; municipal fixed effects were used in both models but are not shown in the presentation; cluster robust standard errors (municipal level) in parentheses.
$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

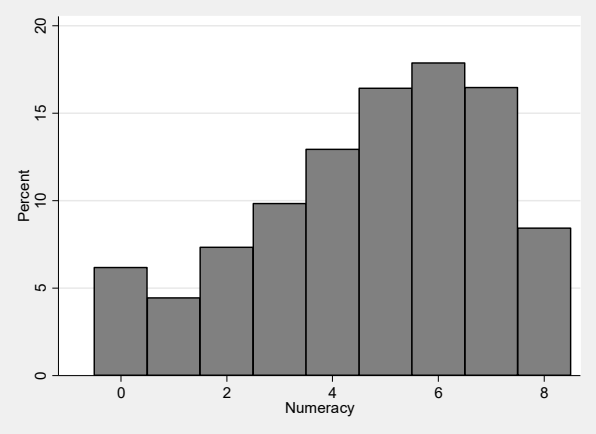
Table A1: Treatments in the ability test

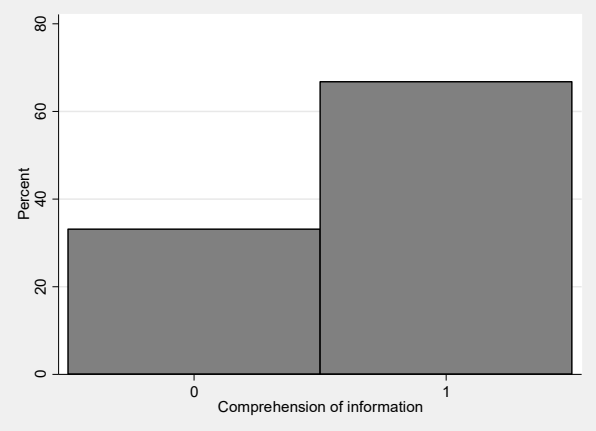
Treatment 1:		
	Operations without complications	Operations with complications
Public hospital	203	68
Private hospital	47	9

Treatment 2:		
	Operations without complications	Operations with complications
Public hospital	68	203
Private hospital	9	47

Note: In both treatments, “Hospital A” and “Hospital B” were used for half of the respondents instead of “Public hospital” and “Private hospital.”

Figure 1: Histograms showing the distribution of the numeracy and dependent variables





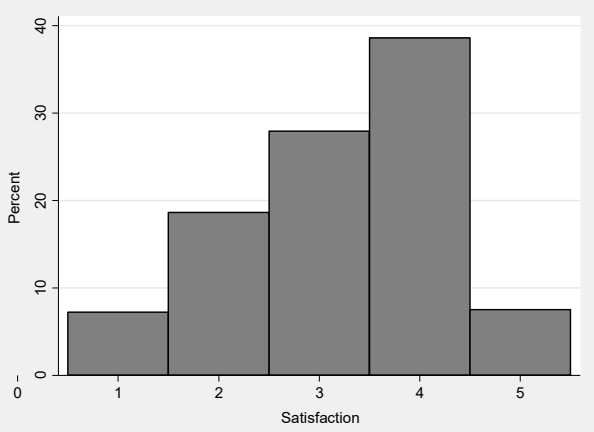
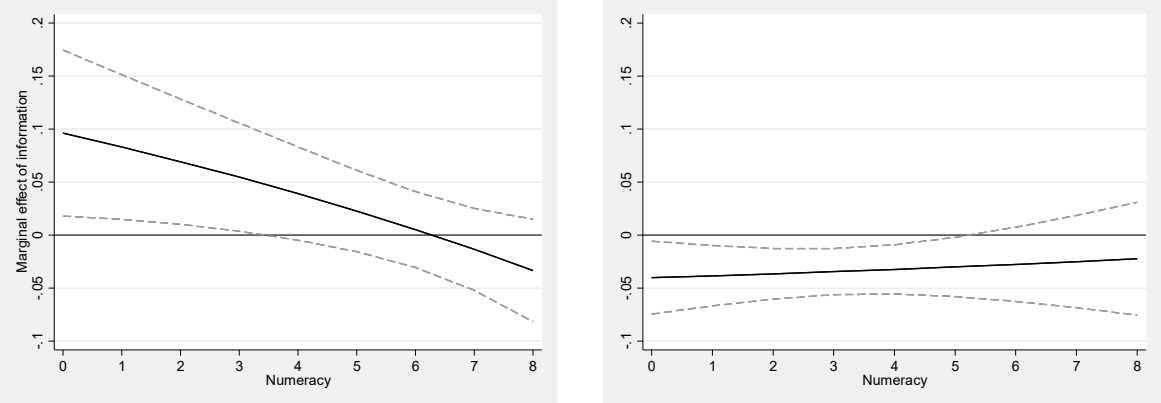


Figure 2: The average marginal effect of information on being very satisfied for different levels of numeracy

Panel A: High performance **Panel B: Low performance**



The figure is based on the analysis in Table 6. 95% confidence intervals shown with dashed lines.