

Research Article

Does performance disclosure affect user satisfaction, voice, and exit? Experimental evidence from service users

Peter Rasmussen Damgaard*, Poul A. Nielsen[†]

Abstract: An emerging literature in behavioral public administration shows that performance information affects the perceptions and choices of citizens vis-a-vis public services and programs. Methodologically, a significant share of these studies relies on hypothetical scenario experiments, or they focus on citizen assessments of broader government entities that citizens have little or no direct interaction with or personal information about. Yet, among actual service users, performance data is only one among many sources of information, potentially limiting its influence. Service users might also engage in motivated reasoning, for instance, by questioning the validity and relevance of inconvenient information about service providers they are otherwise happy with, or whom they are responsible for choosing. In this study, we conducted a survey experiment in the field, offering true performance data to service users, namely parents with children in public schools. We consistently find little or no evidence that performance information affects user satisfaction, intended voice and exit behaviors, incumbency voting, or goal prioritization. These findings question the feasibility of using performance information disclosure to affect the judgments and choices of service users, with potentially important downstream effects on the incentives facing public service providers.

Keywords: Performance information, User satisfaction, Voice, Exit, Experiment

Supplements: [Open materials](#)

Disclosing performance information to citizens and service users has been a core element in increasing government transparency and accountability in recent decades (Barrows, Henderson, Peterson, & West, 2016; James, 2011). Underlying this development is the assumption that performance information can empower citizens and users to make more informed assessments and choices about public service providers, including decisions about satisfaction, voice, and exit (James & Moseley, 2014), and to better hold elected officials accountable (James,

2011). In turn, greater transparency and better-informed citizens are argued to place pressure on service providers and elected officials to be more responsive to the needs and demands of users and citizens, thereby spurring public service improvement (Chingos, Henderson, & West, 2012; Moynihan, 2008).

In recent years, a considerable amount of survey-experimental research has examined how performance information affects the perceptions and choices of citizens and service users. In large part, this body of research finds that citizens and voters respond to performance information in predictable ways, while also noting potential biases in how they assess and respond to said information (e.g., Baekgaard & Serritzlew, 2016; Boyne, James, John, & Petrovsky, 2009; James 2011; James & Moseley, 2014; Jilke, Van Ryzin, G., & Walle, 2016; Olsen, 2017a).

However, these studies rely almost exclusively

*University of Southern Denmark

[†] Aarhus University

Address correspondence to Poul A. Nielsen at (poul@ps.au.dk)

Copyright: © 2020. The authors license this article under the terms of the Creative Commons Attribution 4.0 International License.

on hypothetical scenario experiments, or they focus on citizen assessments of broader government entities or functions that citizens have little or no direct interaction with, or prior personal information about. Hypothetical scenario experiments are attractive because they allow full manipulation of treatments without using deception, make fewer restrictions on sample requirements, and remove noise from other factors that might dominate or interact with performance information. The potential challenge with using hypothetical scenarios, however, is that it excludes all prior information that actual service users have through their knowledge of and interaction with specific service providers (Barrows et al., 2016).

Moreover, whereas citizen responses to performance information have been studied extensively, the users of public services have received less attention in experimental research.¹ This distinction is important because the roles of citizens in general and service users differ concerning the types of assessments and decisions that are relevant to consider and, in turn, what their channels of influence and accountability are (Brown, 2007; Hyde, 1991).² Satisfaction is typically measured among service users, and only service users are able to hold public service providers directly accountable through decisions about exit and voice. These decisions have been critical elements in reforms aimed at empowering users and creating incentives for public service providers to improve (Chingos et al., 2012; Jilke et al., 2016; Le Grand, 2007). By contrast, studies of citizens typically focus on more general assessments of government performance, satisfaction, and other attitudinal outcomes, often within a local political entity (e.g., Baekgaard, 2015; Boyne et al., 2009; Deslatte, 2019; James 2011; James & Moseley, 2014).

Service users also have access to many other sources of information other than standardized performance data. Brown (2007) notes that whereas citizens or the public at large determine their satisfaction based on broader public service outcomes, direct service users also, and perhaps predominantly, base their assessments on their personal experiences and interactions with the service provider. Service users may also differ in their experiences with the same service provider, and so aggregate or averaged service outcomes might matter less to them (Kelly & Swindell, 2002; Lerman & McCabe, 2017). More generally, service users often have access to detailed knowledge and information about the actions of service provid-

ers that often elude citizens, including information about other performance dimensions than those subject to performance measurement (Favero & Meier, 2013). Thus, the prior information of service users may render them indifferent to performance information, or at least they may assign it less weight.³ In addition, because of their personal experiences and direct interaction with service-providing personnel, users may exhibit greater loyalty to (or disapproval of) service providers (Brown, 2007). For instance, when exposed to inconvenient information showing low service performance, users who are otherwise satisfied with the service, or who feel the information questions the merit of their own choice of service provider, may engage in motivated reasoning to deem the new information less valid, reliable, or relevant (Christensen, 2018; James & Van Ryzin, 2017; Nielsen & Moynihan, 2017). Accordingly, while some studies of citizens discuss the implications of their findings in terms of exit and voice decisions, users of specific services differ from the broader public in important respects that may affect how they engage with performance information.

To address these issues, we conducted a survey experiment in the field among actual service users, featuring true and realistic performance information about the users' own service-producing organizations. Specifically, we conducted the experiment among parents with children enrolled in public schools in Denmark ($n=1,185$), who received either absolute, relative, or no performance information. Across model specifications and robustness checks, we find no consistent effects of performance information on user satisfaction, intended voice, or exit behaviors. We also found no effects on incumbency voting or goal prioritization. Moreover, we examined if including a social comparison as a reference point made a difference to the findings, which it did not.

These results suggest that service users may differ in their responses to performance information compared to citizens who respond strongly to similar information treatments (Olsen, 2017a) or compared to samples exposed to hypothetical scenarios. An empirical limitation of the study is that we examine only user responses to performance information and contrast these with previous research on citizens or hypothetical scenario responses to performance information. Future research should conduct a direct within-study comparison of user and non-user responses to performance information.

User and Citizen Responses to Performance Information

Satisfaction, voice, and exit

Experimental methods have been used to study citizen responses to performance information across a range of outcomes. Studies have increasingly examined the underlying processes of how citizen responses to performance information are shaped, for instance, by their prior attitudes and expectations (Baekgaard & Serritzlew, 2016; Marvel, 2016; Jacobsen, R., Snyder, J. W., & Saultz, 2015; Van Ryzin, 2013) and the influence of reference points for comparison (Barrows et al., 2016; Charbonneau & Van Ryzin, 2015; James & Moseley, 2014; Olsen, 2017a). Yet, a general finding across the experimental literature is that performance information affects citizen responses in predictable ways. In the context of the expectancy-disconfirmation model, for instance, Van Ryzin (2013) notes that performance affects satisfaction directly, above and beyond its part of the model's disconfirmation variable (performance – expectations).

In our assessment of whether service users respond differently to performance information than citizens, we focus on a comprehensive set of relevant outcomes to ensure that any differences are not specific to a single outcome (Gerber & Green, 2011). Among these, satisfaction with public services is perhaps the most comparable outcome across users and citizens. Satisfaction measures are collected across a wide range of public services and are increasingly used to inform political and bureaucratic decision-making (Andersen & Hjortskov, 2016). The use of satisfaction measures in public organizations has increased since the 1970s, and much scholarly attention has been paid to the links between service quality and satisfaction measures as well as distinctions between users and non-users (Stipak, 1979; Lyons, W. E., Lowery, D., DeHoog, R. H., 1999; Bouckaert & Walle, 2003). In this literature, the role of direct contact with the service provider has previously been noted (Hero & Durand, 1985; Dinsdale & Marson, 1999). Studies have also shown how alternative attitudinal cues, such as partisanship (Jilke, 2018; Jilke & Baekgaard, 2020) and underlying attitudes toward the public sector (Poister & Henry, 1994; Marvel, 2016) can be sources of important differences between citizens' and users' satisfaction with the same service organization. Except for public services that affect the general population

directly, such as police or sanitation, satisfaction measures are collected primarily from direct service users, which underlines the importance of user satisfaction.

The disclosure of performance information has also been aimed at empowering users to voice their concerns and make more informed decisions about exit and choice among service providers (Dowding & John, 2008; Hirschman, 1970). These types of decisions are generally less relevant to the public at large, although James and Moseley (2014) also examined whether performance information about local government waste recycling affected collective voice behavior among citizens. Voice behavior can consist of complaints or petitions to service providers or elected politicians, but the establishment of formal user boards also offers a channel for users to get more directly involved in influencing how public service organizations operate and prioritize (Torfing, J, Sørensen, E., & Roiseland, 2019). While exit and voice behaviors both are expected to increase when users are exposed to low performance signals, there is also a trade-off between them in the sense that a decision to exit reduces the utility of giving voice (Hirschman, 1970). Whether users choose exit or voice likely depends on the costs and potential benefits of both options, which differs across service contexts. While effects of performance information on user voice and exit have not been studied experimentally, recent quasi-experimental regression-discontinuity studies in the context of No Child Left Behind (NCLB) have found that clear and high-stakes dichotomous school failure signals increased school exit among users (Holbein, 2016; Holbein & Hassell, 2019). These studies also found that failure signals increased citizen voice in terms of voter turnout in school board elections, which, however, are not restricted to service users. These findings indicate that clear failure signals with high-stakes consequences do matter to service users. At the same time, it is an open question whether this finding travels to other contexts without dichotomized failure signals and with less severe consequences tied to performance measures. In the NCLB context, the consequences of failure signals included the risk of school closure. Previous work has additionally demonstrated downstream effects of failure signals, such as falling housing values (Figlio & Lucas 2004; see also Holbein, 2016), that may have separate effects on user responses beyond the direct attitudinal effects of performance information.

Incumbency voting and goal prioritization

We examine two additional outcomes that are less central, but which have received some attention in experimental research. Related to research on retrospective voting, Boyne et al. (2009) provide observational panel evidence that negative performance signals about local governments affect citizens' intentions to vote for the incumbent political party, whereas James (2011) found no incumbency voting effects in an experimental setting despite demonstrated effects on citizen satisfaction. Considering that we focus on users receiving information about specific service providers rather than the broader local government, however, it is less likely that performance information would affect users' voting intention, even if local governments are ultimately charged with funding and organizing service provision.

We also examine whether performance signals affect how users prioritize between different potentially salient organizational goals. Studies of organizational learning from performance feedback have shown that managers prioritize goal dimensions showing poor performance (Holm, 2018; Nielsen, 2014). However, Christensen (2018) found that students exposed to randomized performance signals about their own and a rival university assigned less importance to poorly performing goal dimensions, consistent with the notion that users engage in motivated reasoning about performance data to defend their choice of service provider.

Finally, across these five outcomes, we generally expect that including a social reference point for comparison will increase the effect of performance information (Barrows et al., 2016; Charbonneau & Van Ryzin, 2015; James & Moseley, 2014; Olsen, 2017a). Another important finding to bear in mind is that negative performance information is often found to elicit stronger responses than similarly positive information (James, 2011; Nielsen & Moynihan, 2017; Olsen, 2017a).

Research design and data

The study was designed to achieve a high degree of realism by studying how actual users respond to realistic and truthful performance information about

their own organizations. We conducted the study in the context of public education in Denmark, where standardized performance information is highly comparable and of relatively high quality compared to other types of public services and where the information is made publicly available through government websites. It is also a setting where parents show a high degree of interest and where they can choose between public schools within the same municipality, provided that alternative schools can take in additional students. Parents are also free to choose among private non-profit school alternatives, which are funded through a voucher scheme with a limited co-payment of on average \$1,900 per year (DKR 12,659) in 2017. While schools are governed by municipalities, parents are also represented through school boards, where they can give voice to their concerns and influence school priorities. Public schools typically cover grades 0-9 until students are 15-16 years old, which means that failing to respond to performance information may have long-term implications. Accordingly, this is a setting where we would expect users to be responsive to performance information.

Experimental design

To examine the impact of performance information, we conducted a survey experiment in the field among parents with children enrolled in public schools. Due to ethical concerns, we did not use deception and therefore did not randomize the content of the information. Instead, we offered respondents either no information or true information about their own school's performance. For the treatments, we used official performance metrics from the Ministry of Education. We designed the two treatments to correspond closely to those used by Olsen (2017a), who also examined absolute and relative performance treatments in Danish public education and found clear information effects in a nationally representative citizen sample asked to make judgments about a hypothetical school. Because we focus on real schools and users, we follow Barrows et al. (2016) in including a control group that received no information.⁴ The experimental design is illustrated in Table 1.

Table 1
Experimental Treatments.

Control group	Treatment groups	
<i>No information</i>	<i>Absolute performance</i>	<i>Relative performance</i>
The Ministry of Education publishes annual performance information about public schools in Denmark.	The Ministry of Education publishes annual performance information about public schools in Denmark.	The Ministry of Education publishes annual performance information about public schools in Denmark.
	Below you will find the latest grades from the website of the Ministry of Education.	Below you will find the latest grades from the website of the Ministry of Education.
	GPA of [Name of School]: x_1	GPA of [Name of School]: x_1
		GPA of [Name of Municipality]: x_2

Notes: x_1 and x_2 varied according to the school and municipality of the individual respondent.

As discussed in prior studies employing related experimental designs to other types of actors, control group participants could have some prior understanding of how well their school was performing (e.g., Nielsen & Baekgaard, 2015). Nevertheless, the explicit performance signals that treated participants are exposed to are likely to increase their confidence in making judgments (Nielsen & Moynihan, 2017). In addition, only the treated participants are primed to explicitly consider performance data (Barrows et al., 2016). Moreover, previous studies among supposedly more knowledgeable actors, such as elected local politicians and school teachers, have identified substantial information effects using similar types of treatments (Geys & Sørensen, 2018; Nielsen & Moynihan, 2017; Petersen, N. B. G., Laumann, T. V., Jakobsen, & M., 2018).

Data and measures

We distributed the surveys by sharing survey links on the schools' internal electronic message boards upon agreement with the school principal, school board, or municipal administration. Out of the 170 schools we contacted, 15 schools granted us access to making the survey available to parents. We obtained valid responses from 1,185 parents.⁵ The data were collected in two waves in 2017 (July-September and October-

November), timed before and after the release of new official performance metrics. We did so in order to examine whether newly updated performance information, which fewer or no parents could have been familiar with, had a stronger impact. We found no indication that the timing affected the findings.

As in other studies (Hjortskov, 2019; James & Moseley, 2014; Roch & Poister, 2006; Van Ryzin, 2004), we measured user satisfaction with a single-item question inquiring "How satisfied are you with your child's school overall?" Because of the institutional setup in Danish public schools with direct user involvement through school boards, we measured voice behavior by asking respondents how likely they were to participate in school board work. We note that this type of voicing behavior is one among multiple types of voice behavior, and that it has a higher participation threshold than alternatives such as filing complaints, but also that it arguably allows a greater influence on the subsequent operation of public services. Methodologically, we expect that the more neutral participation framing enables measurement that is less susceptible to social desirability bias compared to more negatively loaded terms such as complaining. The availability of school choice allowed us to measure intended exit by asking parents how likely they were to move their child to a different school.

To measure goal prioritization, we adopted a rank-order battery asking parents to rank how they would prioritize the school effort among different goal dimensions. We included six options that are all salient in the school context in Denmark. Among these, only two options related specifically to student academic performance, namely “student academic achievement” and “preparation for upper secondary education.” We created an additive index of these two items to measure how parents ranked academic performance relative to other goal dimensions. Table A1 in the online appendix shows all items and response categories, and table A2 provides descriptive statistics and correlations.

Estimation strategy and covariate balance

The focus of the experimental design and subsequent analyses is the comparison of treated and untreated respondents from the same schools. For the analyses, we first categorized the participating schools into high-performing, medium-performing, and low-performing depending on their performance relative to the municipal average. We then compared treated and untreated respondents within these groups. For the main analyses we coded schools as high-/low-performing if they performed at least 0.2 grade points above/below the municipal average. The high performing schools on average performed 0.87 grade points above the municipal average (ranging from 0.2 to 1.6), corresponding to 1.7 standard deviation above the municipal average. The low performing school on average performed 0.29 grade points below the municipal average (ranging from -0.2 to -0.8), corresponding to 0.6 standard deviation below the municipal average. Accordingly, this difference of 2.3 standard deviations in the performance information received by parents in the high and low performing schools is substantial.⁶ The residual medium performance group captures 12-20% of schools in each municipality.

The samples in the high- and low-performing scenarios would have been somewhat larger if all schools above/below the municipal average had been categorized as high-/low-performing. However, this would come at the cost of the treatment signal, as respondents from schools placed, for instance, just below the

municipal average might not consider this low performance. Because this cut-off is arbitrary, we also conducted sensitivity analyses with cut-offs at every 0.05 increment starting from the municipal average. We also analyzed the data without making assumptions about discrete high- and low-performance signals by interacting the treatment status with the actual absolute and relative performance levels of the schools.

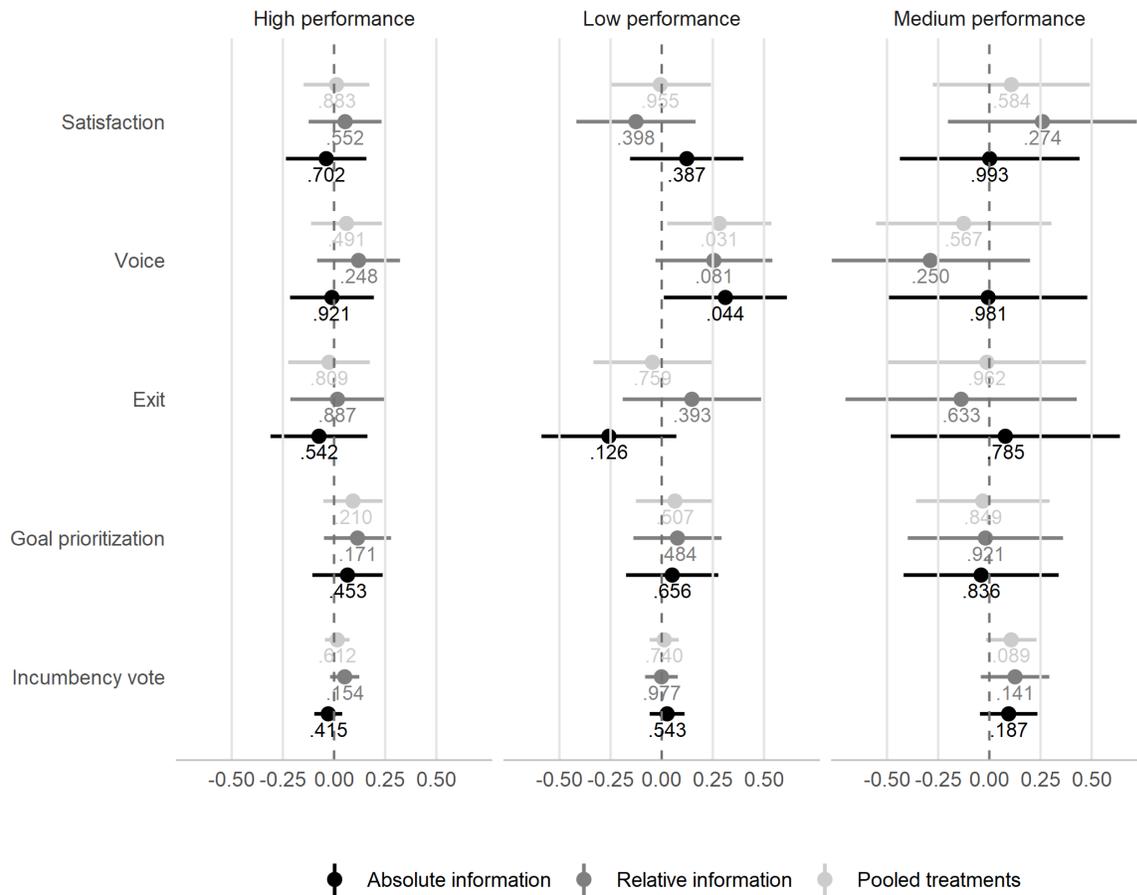
Prior to the experimental conditions, respondents were exposed to general questions about background characteristics (age, years with children at the school, gender, and education level). The only unbalanced covariate across the treatment groups was that among low-performing schools, respondents in the two treatment groups were 1.5 years older than in the control group (see online appendix tables B1 and B2). We return to this imbalance in the findings, as age is also positively correlated with voice and goal prioritization. We found no indications that treatment status affected attrition.

Results

The main findings are presented in figure 1, which illustrates the estimated effects of receiving the absolute and relative performance information treatments (relative to the control group) in the high-, low-, and medium-performance scenarios and for each of the five outcomes. The figure also displays a combined estimate where the two treatment groups are pooled together. The underlying OLS models are presented in online appendix table C1.

Across the five outcomes, the general pattern for the absolute and relative performance information treatment cues in the high- and low-performance scenarios (20 estimates in total) is that performance information does not seem to affect the outcomes. Of particular note, we found no indications that any of the performance treatments affected user satisfaction. The treatment estimates are close to null and far from statistically significant. This finding is especially noteworthy because satisfaction levels are generally considered to be directly responsive to performance information and because satisfaction assessments require no additional commitment or costs for users.

Figure 1
Effects of Performance Treatments with 95% Confidence Intervals and P-values.



Notes: OLS-coefficients with 95% confidence intervals for dummy variable indicating treatment of either absolute information or relative information. In both cases, the reference group is the control group, which received no information. The estimate reported for incumbency voting is the marginal effect from a logit regression.

We also found no effects for exit intentions, which could reflect a general lack of responsiveness to performance information, although it might also be linked to the additional costs for families related to changing schools. We found no effects on incumbency voting. Previous studies of incumbency voting tend to focus on broader local government outcomes that are perhaps more readily attributable to elected politicians, whereas local school performance is more removed from local government politicians. Concerning goal priorities, there are no indications that parents started considering academic achievement more important when informed that academic performance was poor. However, there are no indica-

tions either that parents engaged in motivated reasoning by reordering their goal priorities to make them consistent with the performance data. Instead, parents appear to be unresponsive to the data.

The only exception to these null findings is that respondents in the low absolute performance treatment group appear more likely to engage in voicing behavior. However, as previously mentioned, there was a slight age imbalance between the treated and untreated respondents in the low performance group. When we include age as a covariate, the estimates for absolute and relative performance become smaller and statistically insignificant (respectively, $p=.100$ and $p=.145$) (online appendix table C2).

This suggests that the voice-effect is not robust and could be driven by covariate imbalance.⁷

Across the models, we find no systematic differences between the absolute and relative performance treatments. This is surprising considering that performance information has been found to elicit stronger responses when a social reference point is included for comparison. This does not imply that social reference points are unimportant; however, it is suggestive that if other information or personal experiences are more important to service users than standardized performance information, including a social reference point will not necessarily make a difference.

To further assess the robustness of the findings, we conducted a number of sensitivity analyses. First, we re-analyzed the data using different estimators, including school fixed effects (online appendix table C3) and ordered logistic regression (table D1). Again, we find no significant treatment effects although the estimates for voice are similar to those presented in figure 1. After controlling for age, the estimates again become smaller and statistically insignificant with *p*-values above 0.1. We also conducted non-parametric Mann Whitney-U/Wilcox tests that make less restrictive assumptions, yielding results consistent with those reported in figure 1 (table E1).

Second, we examined whether the findings were sensitive to the placement of the threshold for dividing schools into the low- and high-performance categories (figures F1-F8). For all thresholds higher than 0.20, all estimates become statistically insignificant. The low performance finding for voice is more robust at thresholds lower than 0.20, but again after controlling for age, the treatment effects for voice become insignificant.

Third, we analyzed the data without making assumptions about discrete high- and low-performance signals by interacting the treatment status with the absolute and relative performance levels. A cross-sectional correlation between performance and satisfaction could be endogenous, for instance, because of underlying characteristics of the schools or parents that affect both performance and satisfaction. By contrast, the logic of this moderation approach is that if the observed correlations between performance and satisfaction differ between the treatment and control groups, this difference in correlations can be attributed to the exogenous treatments. As shown in online appendix table G1, we find no evidence of significant interactions across any of the outcomes, in-

cluding for voice where both the absolute and relative performance treatments are far from statistically significant (respectively, $p=.680$ and $p=.255$).⁸

To assess whether these null findings provide support for the absence of meaningful effects, we conducted equivalence tests that examine whether the estimates are significantly lower than pre-specified bounds of effects sizes of interest (Lakens, 2017). Using two one-sided Welch's *t*-tests, we first examined whether the estimated effects were significantly lower than a medium-sized Cohen's *d* of 0.5.⁹ As shown in online appendix table H1, all effect sizes are significantly lower than a medium effect size, suggesting that information effects are at most limited. We also examined whether the findings are significantly lower than a small effect size, specified as $d<0.25$ (online appendix table H2). The findings across the high-performance treatment estimates tend to be significantly lower than this bound, but we cannot rule out that the low performance treatments could have small effects. Yet, it is important to note that this only speaks to the uncertainty of the findings and should not be interpreted as a presence of small effects. It should also be noted that the equivalence tests were conducted for the main model presented in figure 1, which was a best-case scenario in the sense that correcting for covariate imbalance and conducting various sensitivity analyses resulted in weaker and more uncertain effect estimates.

Discussion and Conclusion

Service users have received little attention in experimental research on performance information even though only service users can hold public service providers directly accountable through decisions about exit and voice. Whereas prior research has found that citizens respond to performance information in predictable, albeit biased ways, we find that direct service users were generally unresponsive to performance information across a range of outcomes. Perhaps particularly surprising, we consistently found no influence of performance information on user satisfaction, which is a measure that, unlike exit and voice decisions, entails no costs or commitments for users. These findings suggest that we cannot infer from studies focused on citizens or using hypothetical scenario vignettes how service users in practice will respond to performance information. Service users might still care about service outcomes, but the

results indicate that service users may focus on other information sources than standardized performance information when forming their attitudes. Service users might also engage in motivated reasoning to discount the relevance or validity of performance information that contradicts their prior beliefs.

We examined a context, public education in Denmark, where performance outcomes are salient to school parents and where exit, choice, and voice options are available, which should be favorable conditions for the influence of performance information. These findings therefore question the feasibility of using performance information disclosure to empower service users and inform their assessments and choices among service providers. In turn, this could potentially weaken the incentives facing public service providers to improve performance and be responsive to the needs and demands of service users, especially if information disclosure and choice reforms have weakened traditional hierarchical oversight and accountability mechanisms. A potential limitation of this study is the sampling process, which only includes parents who chose to answer the survey, but the direction of potential bias for out-of-sample inference is not obvious. For instance, answering the survey can reflect both a high level of engagement and satisfaction with the school, which may entail less responsiveness to treatments, or a low level of prior satisfaction, in which case answering the survey is itself a type of voicing. Future work should consider replicating this experimental design using a more precisely defined sample of users.

Another empirical limitation is that we do not sample, and therefore cannot compare directly against, non-users. Instead, our inferences are based on a comparison between our sample of users and previous research using hypothetical vignette experiments, broader citizen samples, or less precisely defined samples which might include both users and non-users. To partially address this, we used experimental treatment formulations adopted from a study of citizens placed in the same national school context (Olsen, 2017a). Nevertheless, future research should seek to directly compare user and non-user responses within the same experimental study.

Future research should seek to replicate and extend the study of user responses to performance information. It would be important to address whether the service context, for instance, the quality of performance data or the availability of choice and voice options, affects user responses. Similarly, the charac-

teristics of users, such as the frequency and intensity of their interactions with the service provider, or whether they are customers, clients, or captives (Brown, 2007), could play an important role. In some cases, it can also be difficult to distinguish clearly between the roles of citizens and service users, for instance, concerning local government utility provision (James & Moseley, 2014), which moreover has been described as a low-information environment (Barrows et al., 2016). Another central question is when users stop responding to performance information. Among future users, Hastings and Weinstein (2008) found that parents chose better-performing schools when they received performance information. This is consistent with the notion that pre-choice users are more likely to incorporate new information, whereas post-choice users engage in motivated reasoning to defend their initial choice (Christensen, 2018). However, it could also be explained by present users (rationally) incorporating their prior information and personal experiences.

Finally, as previously mentioned Holbein (2016) and Holbein & Hassell (2019) found that users were responsive to performance failure signals in the context of NCLB. This quasi-experimental finding could be interpreted as contradictory to the argument and findings in this study. Another interpretation, however, would result in the hypothesis that users become more (less) responsive to performance information as performance signals become more (less) clear and consequential. Future research could consider how to test such a hypothesis, for instance, by building more evidence within either type of context or by seeking to manipulate the clarity and consequences of performance information within a quasi-experimental or survey-experimental setting. Accordingly, future research should explore the underlying mechanisms that affect whether and when service users become responsive to performance information.

Notes

1. Some of these broader citizen samples might also include respondents who are service users, but they are typically not asked to assess their own local service provider. They are instead asked, for instance, how well a broader local government is

performing. In some cases, though, this distinction becomes less clear-cut. For instance, a study by Barrows et al. (2016) exposes a nationally representative sample, which likely includes some school parents, to information about the school district (but not the local school). Generally, the distinction and direct comparison between users and non-users does not seem to have been addressed systematically.

2. Citizens or voters asked to assess broader (local) government performance or, for instance, holding local government incumbents accountable for performance (e.g., Boyne et al. 2009; James 2011) could potentially also be viewed as users of baskets of local government services. In this paper, we apply the term user to refer specifically to direct service users of specific service providers (see also Brown 2007).
3. It is interesting to note that among citizens, episodic frames that describe specific situations have been found to have greater influence on performance evaluations than numerical performance data (Olsen, 2017b)
4. The available survey software did not have a randomization function embedded. Instead, we assigned the participants to treatment and control groups using an exogenous and as-if random factor, namely the participants' date of birth, irrespective of their month and year of birth. Participants born on 11-15th and 26th-31st of any given month or year were assigned to the control group, 6-10th and 21st-25th to treatment group 1, and 1st-5th and 16-20th to treatment group 2.
5. Respondents were debriefed about the experimental nature of the study in a follow-up email. Because the survey was not sent directly to parents, and we are unable to track how many parents actually saw the electronic survey message and link, it is not possible to assess the exact response rate.
6. These calculations are based on the average within-municipality standard deviation of school grade point averages, which is chosen here because parents are likely to primarily compare the performance of their own school to other schools within the same municipality. Because of

between-municipal differences, the standard deviation across all public schools in Denmark is slightly higher. Nevertheless, based on a nationally pooled standard deviation, the performance information received by parents in the high and low performing schools on average still differ by 1.53 standard deviations.

7. We also note that if Bonferroni or other alpha-level corrections for multiple comparisons are applied, none of the estimates are close to being statistically significant.
8. We also examined whether parents were more responsive to performance information if they had less experience with their school. This was measured by the number of years they had had a child enrolled. As shown in online appendix tables I1 and I2, we generally found no differences in treatment effects between users with less than one year or less than two years of experience with the school compared to parents with more experience. Among the 20 interaction estimates, the only exception was that parents with children enrolled for less than one year were significantly less likely to consider exiting when exposed to a high relative performance cue compared to parents with longer experience. However, the sample sizes for this subgroup were only 23 and 24, respectively, for the treatment and control groups, which questions the validity of these findings as small sample sizes are known to increase the risk of false positives (Loken & Gelman, 2017). A linear interaction model with time as a continuous variable showed no indication of differences in treatment effects. While this suggests that performance information has little influence on service users even when they have relatively limited experience with the service provider, it should be noted that we are unable to identify treatment effects among users with more short-term experience.
9. Effect sizes in experimental studies of citizens differ, so there is no clear benchmark against which to compare the effects. While the experimental setups differ, we study the same context and use similar treatments as Olsen (2017a), who reported large effect sizes equal to or greater than an r of almost 0.5, corresponding to a Cohen's d larger than 1.

References

- Andersen, S. C., & Hjortskov, M. (2016). Cognitive biases in performance evaluations. *Journal of Public Administration Research and Theory*, 26(4): 647–62.
- Baekgaard, M. (2015). Performance information and citizen service attitudes: Do cost information and service use affect the relationship? *International Public Management Journal*, 18(2), 228–245.
- Baekgaard, M., & Serritzlew, S. (2016). Interpreting performance information: Motivated reasoning or unbiased comprehension. *Public Administration Review*, 76(1): 73–82.
- Barrows, S., Henderson, M., Peterson, P. E., & West, M. R. (2016). Relative performance information and perceptions of public service quality: Evidence from American school districts. *Journal of Public Administration Research and Theory*, 26(3), 571–583.
- Bouckaert, G., & Van de Walle, S. (2003). Comparing measures of citizen trust and user satisfaction as indicators of ‘good governance’: Difficulties in linking trust and satisfaction indicators. *International Review of Administrative Sciences*, 69(3), 329–343.
- Boyne, G. A., O. James, P. John, and N. Petrovsky. 2009. Democracy and government performance: Holding incumbents accountable in English local governments. *Journal of Politics*, 71(4):1273–84.
- Brown, T. (2007). Coercion versus choice: Citizen evaluations of public service quality across methods of consumption. *Public Administration Review*, 67(3): 559–72.
- Charbonneau, É., & Van Ryzin, G. G. (2015). Benchmarks and citizen judgments of local government performance: Findings from a survey experiment. *Public Management Review*, 17(2), 288–304.
- 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.
- Chingos, M. M., Henderson, M. & West, M. R. (2012). Citizen perceptions of government service quality: Evidence from public schools. *Quarterly Journal of Political Science*, 7(4), 411–445.
- Deslatte, A. (2019). A Bayesian approach for behavioral public administration: Citizen assessments of local government sustainability performance. *Journal of Behavioral Public Administration*, 2(1), 1–12
- Dinsdale, Geoff and Marson, D. Brian (1999) *Citizen/client surveys: Dispelling myths and redrawing maps*. Ottawa: Canadian Centre for Management Development
- Dowding, K., & John, P. (2008). The three exit, three voice and loyalty framework: A test with survey data on local services. *Political Studies*, 56(2), 288–311.
- Favero, N., & Meier, K. J. (2013). Evaluating urban public schools: Parents, teachers, and state assessments. *Public Administration Review*, 73(3), 401–412.
- Figlio, D. N., & Lucas, M. E. (2004). What’s in a grade? School report cards and the housing market. *American Economic Review*, 94(3), 591–604.
- Gerber, A. S., & Green, D. P. (2011). Field experiments and natural experiments. *The Oxford Handbook of Political Science*, July.
- Geys, B., & Sørensen, R. J. (2018). Never change a winning policy? Public sector performance and politicians’ preferences for reforms. *Public Administration Review*, 78 (2), 206–16.
- Hastings, J. S., & Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly Journal of Economics*, 123(4), 1373–1414.
- Hero, R. E., & Durand, R. (1985). Explaining citizen evaluations of urban services: A comparison of some alternative models. *Urban Affairs Quarterly*, 20(3), 344–354.
- Hirschman, A. O. (1970). *Exit, voice, loyalty: Responses to decline in firms, organizations, and states*. Cambridge, MA: Harvard University Press.
- Holbein, J. (2016). Left behind? Citizen responsiveness to government performance information. *American Political Science Review*, 110(2), 353–368.
- Holbein, J. B., & Hassell, H. J. G. (2019). When your group fails: The effect of race-based performance signals on citizen voice and exit. *Journal of Public Administration Research and Theory*, 29(2), 268–286.
- Holm, J. M. (2018). Successful problem solvers? Managerial performance information use to improve low organizational performance. *Journal of Public Administration Research and Theory*, 28(3), 303–320.
- Hjortskov, M. (2019). Citizen expectations and satisfaction over time: Findings from a large sample panel survey of public school parents in Denmark. *The American Review of Public Administration*, 49(3), 353–71.
- Hyde, A. C. (1991). Feedback from customers, clients, and captives. *The Bureaucrat*, 20(4), 49–53.
- Jacobsen, R., Snyder, J. W., & Saultz, A. (2015). Understanding satisfaction with Schools: The role of expectations. *Journal of Public Administration Research and Theory*, 25(3), 831–48.
- James, O. (2011). Performance measures and democracy: Information effects on citizens in field and laboratory experiments. *Journal of Public Administration Research and Theory*, 21(3), 399–418.
- James, O., & Moseley, A. (2014). Does performance information about public services affect citizens’ perceptions, satisfaction, and voice behaviour? Field experiments with absolute and relative performance information. *Public Administration*, 92(2), 493–511.
- James, O., & Van Ryzin, G. G. (2017). Motivated

- reasoning about public performance: An experimental study of how citizens judge the affordable care act. *Journal of Public Administration Research and Theory*, 27(1), 197–209.
- Jilke, S. (2018). Citizen satisfaction under changing political leadership: The role of partisan motivated reasoning. *Governance*, 31(3), 515–533.
- Jilke, S., & Baekgaard, M. (2020). The political psychology of citizen satisfaction: Does functional responsibility matter? *Journal of Public Administration Research and Theory*, 30(1), 130–143.
- Jilke, S., Van Ryzin, G. G., & Van de Walle, S. (2016). Responses to decline in marketized public services: An experimental evaluation of choice overload. *Journal of Public Administration Research and Theory*, 26(3), 421–32.
- Le Grand, J. (2007). *The other invisible hand: Delivering public services through choice and competition*. Princeton, Oxford: Princeton University Press.
- Kelly, J. M., & Swindell, D. (2002). A multiple-indicator approach to municipal service evaluation: Correlating performance measurement and citizen satisfaction across jurisdictions. *Public Administration Review*, 62(5), 610–21.
- Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355–62.
- Lerman, A. E., and McCabe, K. T. (2017). Personal experience and public opinion: A theory and test of conditional policy feedback. *The Journal of Politics*, 79(2), 624–41.
- Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, 355(6325), 584–585.
- Lyons, W. E., Lowery, D., DeHoog, R. H. (1999). *The politics of dissatisfaction: Citizens, services and urban institutions*. Armonk, New York: M. E. Sharpe, Inc.
- Marvel, J. D. (2016). Unconscious bias in citizens' evaluations of public sector performance. *Journal of Public Administration Research and Theory*, 26(1), 143–58.
- Moynihan, D. P. (2008). *The dynamics of performance management: Constructing information and reform*. Washington, DC: Georgetown University Press.
- Nielsen, P. A. (2014). Learning from performance feedback: Performance information, aspiration levels, and managerial priorities. *Public Administration*, 92(1), 142–160.
- Nielsen, P. A., & Baekgaard, M. (2015). Performance information, blame avoidance, and politicians' attitudes to spending and reform: Evidence from an experiment. *Journal of Public Administration Research and Theory*, 25(2), 545–569.
- Nielsen, P. A., & Moynihan, D. P. (2017). How do politicians attribute bureaucratic responsibility for performance? Negativity bias and interest group advocacy. *Journal of Public Administration Research and Theory*, 27(2), 269–283.
- Olsen, A. L. (2017a). Compared to what? How social and historical reference points affect citizens' Performance evaluations. *Journal of Public Administration Research and Theory*, 27(4), 562–580.
- Olsen, A. L. (2017b). Human Interest or Hard Numbers? Experiments on Citizens' Selection, Exposure, and Recall of Performance Information. *Public Administration Review*, 77(3), 408–420.
- Petersen, N. B. G., Laumann, T. V., & Jakobsen, M. (2018). Acceptance or disapproval: Performance information in the eyes of public frontline employees. *Journal of Public Administration Research and Theory*, 29(1), 101–117.
- Poister, T. H., & Henry, G. T. (1994). Citizen ratings of public and private service quality: A comparative perspective. *Public Administration Review*, 54(2), 155–160.
- Roch, C. H., & Poister, T. H. (2006). Citizens, accountability, and service satisfaction: The influence of expectations. *Urban Affairs Review*, 41(3), 292–308.
- Van Ryzin, G. G. (2004). Expectations, performance, and citizen satisfaction with urban services. *Journal of Policy Analysis and Management*, 23(3), 433–448.
- Van Ryzin, G. G. (2013). An experimental test of the expectancy-disconfirmation theory of citizen satisfaction. *Journal of Policy Analysis and Management*, 32(3), 597–614.
- Stipak, B. (1979). Citizen satisfaction with urban services: Potential misuse as a performance indicator. *Public Administration Review*, 39(1), 46–52.
- Torfinng, J., Sørensen, E. & Røiseland, A. (2019). Transforming the public sector into an arena for co-creation: Barriers, drivers, benefits, and ways forward. *Administration & Society*, 51(5), 795–825.