

**Making sense of performance information on effectiveness,
costs, and equality during the COVID-19 pandemic.
Differences across joint and separate evaluations**

Author(s) and affiliation(s)

Maria Falk Mikkelsen*
VIVE – The Danish Research Center for Social Research

Address correspondence to Maria Falk Mikkelsen at mfm@vive.dk

This article has been accepted for publication but has not yet been through the production process (formatting, copy-editing, issuing of Open Science badges, if applicable). Please cite this article as doi: 10.30636/jbpa.42.205

Making sense of performance information on effectiveness, costs, and equality during the COVID-19 pandemic. Differences across joint and separate evaluations

Abstract:

This paper uses the COVID-19 pandemic as an extreme case to test differences in citizens' use of performance information on effectiveness, cost, and equality in separate (without a reference point) and joint evaluations (with a reference point). Using a pre-registered between-subject conjoint survey experiment on 2,025 Danish citizens from early spring 2020, the paper tests how citizens rate government strategies to combat COVID-19. The strategies varied in terms of effectiveness (mortality rate), costs (overall economic costs), and equality (distribution of economic costs and access to testing). Respondents were randomly drawn to rate either one strategy—with no opportunity to compare performance information between strategies— or two strategies. Drawing on the evaluability hypothesis, the paper argues and shows that in the separate evaluations only performance information on equality significantly affects citizens' ratings. While citizens disregard harder-to-interpret performance information on effectiveness and cost, the relational connotations of equality measures aid citizens' interpretation of such data even without reference points. In joint evaluations, however, performance information on mortality rate and economic cost affects citizens' ratings. These results suggest that easier-to-interpret—but not necessarily more important—performance information can dominate citizen decision-making when no 'yardstick' is available.

Keywords: COVID-19, Performance information use, Survey experiments, Reference points

With the outbreak of the COVID-19 pandemic, citizens around the globe have for months received daily updates on their governments' efforts to fight the COVID-19, for example information on mortality rates, the number of newly infected, access to testing, overall costs, and the number of vaccinated citizens. Providing citizens with performance information on public service delivery is widely viewed as important for democratic accountability (Behn, 2003; Moynihan, 2008; van Dooren et al., 2015; van Dooren & van de Walle, 2008). Performance information is expected to enlighten citizens and potentially trigger voice and/or exit (Boyne et al., 2002; Hirschman, 1970). However, how do citizens make sense of and use such performance information? In particular, given exposure to multiple types of performance information, what information do citizens draw upon when making decisions about services?

While citizens' performance information use has received increasing scholarly attention in recent years (e.g., Baekgaard & Serritzlew, 2016; James, 2011a, 2011b; James & Moseley, 2014; Walker et al., 2018), our knowledge of how citizens use and balance different types of performance information remains scarce. In response, scholars have called attention to the interplay between different types of performance information (James & Moseley, 2014; Walker et al., 2018).

This paper contributes to the literature by testing whether citizens' use of performance information on effectiveness, cost, and equality in decision-making differs in separate (without a reference point) and joint (with a reference point) evaluations. Previous research shows that a relevant 'yardstick' or a reference point can aid citizens' interpretation of

performance information on effectiveness (e.g., student performance, rate of waste recycling, municipal performance, and unemployment rates) (Charbonneau & Van Ryzin, 2015; James, 2011a, 2011b; James & Moseley, 2014; Olsen, 2017). Less attention has been paid to how citizens use and balance different types of performance information in separate and joint evaluations.

Drawing on the evaluability hypothesis (Hsee, 1996, 1998; Hsee et al., 1999), this paper argues that in separate evaluations performance information on equality will have a larger effect on citizen decision-making than performance information on effectiveness and cost. To economize on cognitive ability, citizens are expected to use information cues from the most accessible information and neglect those types of information that appear harder to interpret (Bazerman et al., 1992, 1999; Hsee, 1996; Hsee et al., 1999; Kahneman, 2003).

Given the relational connotations of equality measures (capturing differences in access to or levels of public service between groups), such data will convey knowledge of scale properties and is expected to be 'easier' to interpret compared to performance information on effectiveness and cost.

The paper uses the COVID-19 pandemic as an extreme case to test differences in citizens' rating of government strategies to combat COVID-19 in separate and joint evaluations. As holding elected officials and public employees accountable is particularly important during crises, this case provides unique insights into citizens' performance information use in a high-stakes situation.

Theory

Only a few studies within public administration have investigated citizen decision-making when multiple types of performance information are available. Grosso et al. (2017) find that providing both cost and output performance information increased citizen support for spending on an HIV/AIDS program compared to providing output information only. Baekgaard (2015) reports similar findings when combining performance information on outcomes with information on costs, while Walker et al. (2018) find that combining information on equality with information on effectiveness lead to higher ratings of public service. None of these studies, however, investigate the importance of reference points for citizen decision-making.

In contrast, the fields of psychology and economics have paid considerable attention to decision-making in separate and joint evaluations (e.g., Bazerman et al., 1992, 1999; Kahneman, 2003). The evaluability hypothesis suggests that in separate evaluations (when decisions are made without being able to compare between alternatives), decisions will be influenced more by information that is 'easy' to interpret than by information that is 'hard' to interpret. To economize on cognitive ability, citizens use information cues from the most accessible information, even if the hard-to-evaluate information is generally considered more important (Hsee, 1996, 1998; Hsee et al., 1999). While this hypothesis has been confirmed in a number of papers (e.g., Bazerman et al., 1999; Hsee, 1996, 1998; Hsee et al., 1999; List, 2002), application within public administration research has so far been limited.

Information is considered 'hard' to evaluate if the decision maker is not aware of the possible 'distribution' such as the possible 'range' and 'scale' as well as the 'natural' reference point for the information (Hsee, 1998). Under such circumstances, the decision maker will not know whether a given value is 'good' or 'bad'. In contrast, when the cues from the information are clearly related to a 'good' or 'bad' value, information is considered easier to interpret. For example, Hsee (1998) found that in separate evaluations, a dinner set with 24 intact pieces was preferred to a dinner set with 40 pieces (including the 24 intact pieces, 7 other intact pieces and 9 broken pieces). As the respondents in the separate evaluation were not able to compare the number of pieces, they based their decision on the information cues of some pieces being broken. 'Broken' has clear negative connotations, thus the option with broken pieces (while clearly superior) was in separate evaluations rated lower than the option with 24 intact pieces¹.

Interpreting performance information

This paper distinguishes between performance information on effectiveness, cost, and equality. As public performance is a multidimensional construct (e.g., Ostrom, 1973; Rainey, 2009; Walker et al., 2010), this distinction does not capture all relevant dimensions of public performance. However, it does capture an element considered important in most definitions of public performance, namely the difference or potential trade-off between the goal of optimizing production and the goal of pursuing equality and fairness in access to and level of public service (Boyne et al., 2003; Hood, 1991; Selden & Sowa, 2004; Walker et al., 2010). Performance information on effectiveness, cost, and equality thus represents an important source of information on public performance for citizens.

Effectiveness and cost

Performance information on effectiveness and cost captures the extent to which public production delivers ‘high value for money’ with ‘no waste’ (Hood, 1991). However, when a citizen is presented with a value of either cost or effectiveness, it is not self-evident whether such performance constitutes a ‘waste of money’ or ‘high effectiveness’. Another potential source of performance information ambiguity is the lack of a clear causal relationship between the actions of public actors and performance outcomes. Effectiveness and costs are not solely determined by public actors. All principals have trouble determining whether bad outcomes can be blamed on agents or other factors (Miller & Whitford, 2007; Moe, 1984). Citizens—as the ultimate principal—experience the same difficulties. Research finds that reference points aid citizens’ interpretation of performance information on effectiveness (Charbonneau & Van Ryzin, 2015; James, 2011a, 2011b; James & Moseley, 2014; Olsen, 2017). Reference points aid interpretation by providing a yardstick or a potential ‘counterfactual’ (Angrist & Pischke, 2009) against which citizens can measure performance information on effectiveness and cost.

Equality

Performance information on equality captures the extent to which access to and level of public service are equal regardless of such factors as sex, gender, race, and income. As other research, this paper understands equality as ‘equality of opportunity’—providing equal services to different groups—not ‘equality of outcomes’—providing different services to secure equal outcomes for different groups—which is better captured by the concept of

equity (Bronfenbrenner, 1973). As governments have the power to redistribute costs and allocate services, unequal access to and unequal levels of services between groups are more closely tied to the actions of public actors, which may aid interpretation of performance information.

Furthermore, equality is by definition a relational concept. Even if not always clearly stated, equality measures differences in access to or level of public service between groups. For example, if only seriously ill patients are tested for COVID-19, one would—without any more information—also know that non-seriously ill patients will not be tested. The relational connotations of equality measures thus convey some knowledge of scale properties even without reference points.

These relational connotations of equality measures can aid interpretations of whether something is considered ‘fair’ or ‘unfair’. A 30/70 divide of USD 100 will likely be considered unequal regardless of whether an alternative of a 50/50 divide was presented or not. Consequently, experimental studies find that despite individual and cultural differences, participants have a clear view of what constitutes an ‘equal’ or ‘unequal’ divide of a cash prize, even in one-shot games where no reference points are available (Engel, 2011; Fehr & Gächter, 2011; Hoffman et al., 2008) and in more complicated scenarios including efforts (e.g., Franco-Watkins et al., 2013). These experimental studies find that participants are willing to go quite far to punish unfairness. Unfairness can thus provoke a strong emotional response from citizens.

Hypotheses

The above arguments suggest that performance information on cost and effectiveness will be harder for citizens to interpret than performance information on equality in separate evaluations. Due to the relational connotations inherent to measures of equality, such performance information will appear easier to interpret. Drawing on the evaluability hypothesis, the paper thus expects that

H₁: In separate evaluations, performance data on equality will have a larger effect on citizen decision-making than performance information on effectiveness and cost.

H₂: In joint evaluations, performance data on effectiveness and cost will have a larger effect on citizen decision-making than in separate evaluations.

While performance information on equality may appear easier to interpret, it does not follow that performance information on equality will have the same effect on citizen decision-making in separate and joint evaluations. Citizens are not necessarily better at interpreting performance data on equality compared to performance information on effectiveness and cost. Other studies of the evaluability hypothesis clearly show that citizens can make illogical choices in separate evaluations, because they tend to focus on (potentially) unimportant details that they believe they can interpret (but might not be able to). Given the emotional responses ‘unfairness’ has sparked in previous studies, citizens may ‘overreact’ to performance information on equality in separate evaluations.

The importance placed on performance information on effectiveness, cost, and equality when evaluating public performance will likely also differ between citizens. Some citizens and some countries will accept more inequality than others will, and for some types of public services performance information on effectiveness and costs is likely more important for citizens than for other services. Such *level* differences in the overall importance placed on performance information on effectiveness, cost, and equality are not expected to matter for the above hypothesized *relational* differences between separate and joint evaluations. To test the robustness of the results, the paper performs subgroup analysis of right- and left-wing voters, as these two groups likely have different opinions on the importance of performance information on effectiveness, cost, and equality.

Data

The hypotheses were tested using a survey experiment among Danish citizens. The full sample consists of 2,025 responses from an internet panel (Voxmeter) and is representative of the Danish population above 18 years of age. The survey data was collected in 2020 from March 30 to April 6, which was the third week² of a partial lockdown implemented in Denmark as a response to the COVID-19 pandemic.

Experimental design

The survey experiment used a between-subject conjoint experimental design³. The sample was randomly split into two groups (Table A1 in appendix A shows how the two groups balanced on relevant pre-treatment background variables). Both groups were asked to imagine that the government was considering a new strategy to combat COVID-19. One

group, 'Separate evaluation' (1/3 of the sample), were presented with one fictional strategy, and were asked whether they would like the government to pursue this strategy (on a 7-point Likert scale). The other group, 'Joint evaluation' (2/3 of the sample), were presented with two fictional strategies—thus enabling comparisons between the two strategies—and, for each strategy, were asked whether they would like the government to pursue it. Both groups thus evaluated future performance, similar to the participants in Webeck and Nicholson-Crotty's study (2020), and not previous performance.

While the second group rated two strategies, the paper only uses the ratings of the first strategy, thus ensuring that the responses from the two groups are based on the same number of questions⁴. Both groups were made aware that the strategies were *"fictive and thus not representative of actual strategies that Danish authorities may or may not be working on."*

The strategies differed on a total of five different performance indicators: effectiveness (mortality rate), cost (economic cost), equality (access to testing and distribution of cost) and procedural justice (use of private cell data to combat COVID-19) (see table 1), leading to a total of 108 strategies. An example of how the survey experiment was presented to respondents can be found in table A2 in appendix A.

While all models used in this paper control for the variable 'use of cell data', only results for performance information on effectiveness, cost, and equality will be presented in the findings section. Full results are presented in table A3 in appendix A. The order of the

performance information was randomized, and respondents in the ‘Joint evaluation’ were not able to receive two identical strategies.

Table 1. Conjoint design

Performance information on	Dimension	Values
Mortality rate (percentage of infected expected to die from COVID-19)	Effectiveness	0.1-0.3% 0.5-1% 3-5%
Economic cost of COVID-19	Cost	Less than 2008 financial crisis Similar to 2008 financial crisis More than 2008 financial crisis
Access to tests	Equality in access to health care	All with a need Seriously ill Seriously ill and those willing to pay themselves
Distribution of the economic costs	Equality in cost	Equally distributed across groups Low wage-earners and vulnerable groups hit worst
Access private cell information to reduce the spread of COVID-19	Procedural justice	Yes No

Performance information on effectiveness, cost, and equality

Mortality rate and the economic cost are easily the two most important performance outcomes of governments’ COVID-19 responses. Governments are expected to prevent excess deaths and economic losses due to the pandemic.

On March 10, just before the partial lockdown, the Danish health authorities estimated a COVID-19 mortality rate between 0.3 and 1%, which was heavily publicized in the media (Sundhedsstyrelsen, 2020). This estimate is quite close to the estimated mortality of 1.38% published in *The Lancet* on March 30 (Verity et al., 2020). In the survey experiment, respondents were presented with either a mortality rate close to the flu (0.1-0.3%), a mortality rate close to the estimate of the Danish health authorities (0.5-1%) or a higher fatality rate (3-5%)⁵.

Information on economic cost was presented with the text ‘Below, equal to, and above the 2008 financial crisis’ in the experiment. Using text instead of numbers may introduce some ambiguity. However, research has shown that citizens experience particular difficulties in interpreting numbers (Olsen, 2017). Thus, to rule out differences in the interpretation of text and numbers as the main driver of differences in performance information use, the information on cost was presented in text to match the performance information on equality.

The two equality performance indicators capture equality in access to health care (‘Access to testing’) and equality in cost (‘Distribution of the economic costs in relation to COVID-19’). Three values were chosen for access to testing: ‘Everyone in need’, ‘Seriously ill’, and ‘Seriously ill and citizens willing to pay’. One could argue that differences in access to testing between ‘Everyone in need’ and ‘Seriously ill’ may capture legitimate rationalization of health care instead of equality. Thus, the paper will only focus on the differences between ‘Seriously ill’, and ‘Seriously ill and citizens willing to pay’⁶.

If governments do not interfere, the losses due to COVID-19 will likely be quite unevenly distributed across groups. In particular, low wage-earners in vulnerable employment are likely to be hit hard. However, governments have the power to redistribute losses more evenly through compensation schemes. Two values were chosen for the distribution of economic loss: ‘Equally distributed across groups’ and ‘Low wage-earners and vulnerable groups hit worst’⁷.

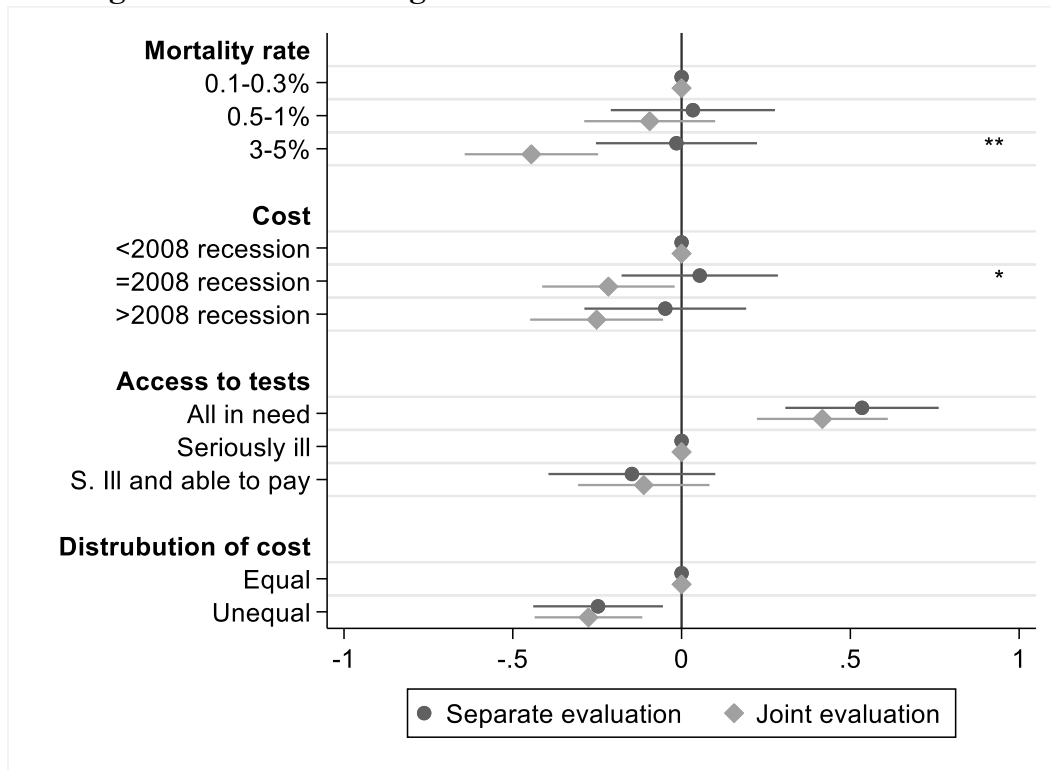
Findings

Marginal effects of performance information on mortality rate, cost, access to tests, and distribution of costs for respondents' ratings were estimated separately for respondents in separate and joint evaluations using linear regression (Hainmueller et al., 2014) (regression results are shown in table A3 in appendix A). Figure 1 displays differences in marginal effects between the separate and joint evaluations.

The results from figure 1 (Separate evaluation) show no significant effect of performance information on effectiveness and cost on respondents' ratings of the COVID-19 strategy; the confidence interval of the estimates of effectiveness and cost contains zero. Thus, when respondents are presented with one strategy, respondents on average do not rate a strategy with a lower mortality rate (0.1-0.3%) as better than a strategy with a higher mortality rate (3-5%), holding other factors constant. Respondents also do not rate a strategy with high economic costs as worse than a strategy with low costs.

Figure 1

The marginal effects of performance information on effectiveness, costs and equality for ratings of COVID-19 strategies



Note: Differences between separate and joint evaluations significant at * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. $N = 2,025$.

In contrast, the two equality performance indicators ('Access to tests' and 'Distribution of cost') both clearly have a significant effect on respondents' rating of the strategy in separate evaluations: more equality leads to higher ratings. Thus, even when no reference points are available, respondents will on average rate a strategy with e.g., an unequal distribution of costs as worse than a strategy with an equal distribution of costs. Strategies in which 'seriously ill patients and those able to pay are tested' are also rated lower than strategies in which 'seriously ill patients are tested'.

The results thus support hypothesis 1: in separate evaluations, performance information on equality will have a larger effect on citizen decision-making than performance information on effectiveness and cost.

Model 2 clearly shows that both the mortality rate and the economic costs have a significant effect on how citizens rate strategies—but only when respondents are able to compare strategies. The estimated difference in respondents' ratings between low and high mortality rate is about 0.4 of a standard deviation in model 2, while the estimated difference between low and high cost is about 0.18 of a standard deviation. These estimates are both significantly different from zero. Figure 1 also shows significant differences between estimates of mortality ('3-5%') and cost ('=2008') between the separate and joint evaluations. Thus, there is evidence to support hypothesis 2: in joint evaluations, performance data on effectiveness and cost has a larger effect on citizen decision-making than in separate evaluations.

Estimates of the equality measures are all smaller for model 2 than model 1. None of these differences are, however, significant. Thus, this paper finds no significant evidence of citizens 'overreacting' to performance information on equality in separate evaluations.

Figure B1 in appendix B shows subgroup analyses for right- and left-wing voters. While performance information on equality is generally less important for ratings by right-wing voters, Figure B1 also shows that both right- and left-wing voters draw exclusively on performance information on equality when rating strategies in separate evaluations. While

results are less strong for right-wing voters (only the equality measure ‘Distribution of cost’ significantly affects ratings), results are still consistent with expectations from hypothesis 1. Figure B1 also shows that differences in the use of performance information on effectiveness between separate and joint evaluations are primarily driven by right-wing voters, while such differences in the use of performance information on cost are primarily driven by left-wing voters.

Discussion and Conclusion

The results from this paper show that citizens’ use of performance information on effectiveness, cost, and equality in decision-making differs in separate and joint evaluations. Using data from a pre-registered between-subject conjoint survey experiment, the paper shows that mortality rates, economic costs, distribution of costs, and access to testing matter for respondents’ ratings of government strategies to combat COVID-19: higher performance (lower mortality rates and costs and higher equality) leads to more favorable ratings in the joint evaluations. The results, however, also show that citizens are unable to utilize performance information on effectiveness and cost in the separate evaluation. When respondents are presented with one strategy—and thus no options to compare information across strategies—only performance information on equality has a significant effect on citizens’ ratings.

These results are consistent with expectations from the evaluability hypothesis (Hsee, 1996, 1998; Hsee & Zhang, 2010). Citizens use information cues from the most accessible

information. Thus, under separate evaluation ‘easier-to-interpret’—but not necessarily more important—performance information dominates citizens’ decision-making.

The paper’s findings have both practical and theoretical implications. As performance information on effectiveness and cost risks being disregarded when performance information on equality or other types of ‘easier-to-interpret’ information is available, providing reference points to citizens is essential for their use of such performance information.

The findings also draw attention to the ‘accessibility’ of different types of performance information. This paper suggests that the relational connotations of performance information on equality aid citizens’ interpretation. However, more work is needed to understand when and under what circumstance performance information is ‘easy’ and ‘hard’ to interpret.

In particular, research on the interpretation of equity and ‘equality of outcomes’ would be valuable. On the one hand, performance information on equity may appear easier to interpret than performance information on effectiveness and cost, as measures of equity—e.g., differences in COVID-19 mortality rates between different groups—would also have relational connotations. On the other hand, public actors are generally less in control of ‘equality of outcomes’ than ‘equality of opportunity’. Differences in mortality rates between groups will not only be the results of (a lack of) public services, but also differences in e.g., lifestyles between groups. Thus, citizens may find it more difficult to assess performance on

equity than equality. Other research finds that performance information on output has a larger significant influence on citizens' evaluation than performance information on outcomes (Grosso et al., 2017). While Grosso et al. do not draw this conclusion, such results could suggest that outputs—more closely related to government action—are easier to interpret than outcomes. However, more theoretical and empirical work is needed before such conclusions can be drawn.

Performance information on equality may play a larger role in the Danish setting than for more liberal and conservative welfare states (Esping-Andersen, 2002) or for more heterogeneous countries. The paper finds that both right- and left-wing voters draw exclusively on performance information on equality when making a decision in the separate evaluation—although performance information on equality is less important for right-wing voters. Such results could suggest that the hypothesized *relational* differences between groups with and without reference points are not affected by *level* differences in the overall importance placed on performance information on effectiveness, cost, and equality. However, new studies in more liberal and conservative welfare states and heterogeneous countries are necessary to confirm such expectations.

As the COVID-19 pandemic is in many ways an extreme case, the uniqueness of the case may place some limitations on the generalizability of the results. As the empirical study was conducted at the beginning of the pandemic, the timing of the study may have made it harder for citizens to interpret performance information in general given that the pandemic was still quite new. In addition, results may not be generalizable to 'normal' public services.

Citizens' performance information use during a 'once-in-a-lifetime' event may differ considerably from performance information use when citizens and service providers have regular interactions (such as e.g., public education).

However, by investigating such an extreme case, novel insight may also be gained. First, holding elected officials and public employees accountable is particularly important during crises. By investigating the COVID-19 pandemic, the paper provides unique insights into citizens' performance information use in a high-stakes situation, and the results may be generalizable to government responses to other crises, e.g., disaster response.

Second, in contrast to previous research, this paper analyzes the government's response to a highly salient crisis with intense media coverage. Knowledge of public performance among the general public is generally low (Baekgaard & Serritzlew, 2016, 2018; Hvidman, 2019; Hvidman & Andersen, 2015). In contrast, public performance relating to COVID-19 (e.g., mortality rates, costs, number of people infected) has received intense media coverage. As the pandemic has life-and-death implications and significantly limits everyday life due to lockdowns, citizens' average interest in public performance related to COVID-19 is likely higher than for any other type of public service. Despite this intense media attention, the results of this study show that respondents still find it difficult to interpret performance information. These results suggest that the interpretation of performance information is inherently difficult even when salience is high.

Acknowledgment

This research was supported by the Carlsberg Foundation.

Notes

- 1) Hsee (1998) present similar findings for a case where respondents were exposed to drawings instead of text and numerical information. Thus, the evaluability hypothesis is also relevant for non-numerical and non-text information.
- 2) During the partial lockdown, which started on March 11, all schools, universities, day care centers, and non-essential health services were closed, and all public employees with non-critical tasks were ordered to work from home. Danes were able to do non-essential shopping, visit family and friends and go on vacation in other parts of the country. However, interaction and movement between regions was strongly discouraged and all gatherings above 10 people in public places were banned. In week three of the partial lockdown, the number of deaths attributed to COVID-19 doubled from a total of 90 to around 180. The total number of cases of COVID19 rose from around 2800 at the beginning of week three to 4600 by the end of week three. Authorities, however, feared that there might have been a much higher number of cases. Due to a lack of tests, only seriously ill patients were tested in week three. The peak in the number of hospitalized COVID-19 patients also occurred during week three (Sundhedsstyrelsen, 2020).
- 3) The between-subject conjoint experimental design was part of a larger experimental design testing the importance of performance information (on effectiveness, economic costs, equality, and procedural justice) for citizens' evaluation of COVID-19 strategies and for trust in the government.
- 4) The results do not change when using the second choice.

- 5) In the experiment, respondents were presented with the mortality rate as 'Percentage of infected expected to die from COVID-19' to avoid terms such as case fatality rate (CFR) or infection fatality rate (IFR), which are likely unfamiliar to some respondents. As the number of diagnosed infected and the total number infected can vary, the experiment contains some ambiguity in relation to whether the mortality rate refers to CFR or IFR; some respondents may have thought of CFR when rating strategies, while others thought of IFR. Such differences will, however, be randomly assigned across the two groups given the experimental design and thus likely do not affect the overall conclusions of the paper.
- 6) Social desirability bias may affect how citizens rate strategies, particularly in relation to the distribution of costs. While social desirability may potentially lead to an overestimation of the effect of 'Distribution of costs' on citizens' ratings of COVID-19 strategies, it does not affect the overall conclusions of the paper. Such an overestimation would have occurred in both the separate and joint evaluations, and thus cannot explain differences in citizens' ratings between models.
- 7) New studies of non-extreme cases should consider operationalizing equality as the difference between free access to services for everyone compared to limited free access for a small group with options to pay for services.

References

- Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.
- 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.
- Baekgaard, M., & Serritzlew, S. (2018). Those who understand it will not be persuaded: A performance information paradox. *International Public Management Journal*, *23*(1), 138-160.
- Bazerman, M. H., Loewenstein, G. F., & White, S. B. (1992). Reversals of preference in allocation decisions: Judging an alternative versus choosing among alternatives. *Administrative Science Quarterly*, *37*(2), 220-240.
- Bazerman, M. H., Moore, D. A., Tenbrunsel, A. E., Wade-Benzoni, K. A., & Blount, S. (1999). Explaining how preferences change across joint versus separate evaluation. *Journal of Economic Behavior and Organization*, *39*(1), 41–58.
- Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review*, *63*(5), 586–606.
- Boyne, G. A., Farrell, C., Law, J., Powell, M., & Walker, R. M. (2003). *Evaluating public management reforms: Principles and practice*. Open University Press.

- Boyne, G., Gould-Williams, J., Law, J., & Walker, R. (2002). Plans, performance information and accountability: The case of Best Value. *Public Administration*, 80(4), 691–710.
- Bronfenbrenner, M. (1973). Equality and equity. *The ANNALS of the American Academy of Political and Social Science*, 409(1), 9–23.
- 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.
- Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4), 583– 610.
- Esping-Andersen, G. (2002). Towards the good society, once again? In G. Esping-Andersen (Ed.), *Why we need a new welfare state* (pp. 1–25). Oxford University Press.
- Fehr, E., & Gächter, S. (2011). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14(3), 159-181.
- Franco-Watkins, A. M., Edwards, B. D., & Acuff, R. E. (2013). Effort and fairness in bargaining games. *Journal of Behavioral Decision Making*, 26(1), 79–90.
- Grosso, A., Charbonneau, É., & Van Ryzin, G. G. (2017). How citizens respond to outputs, outcomes, and costs: A survey experiment about an HIV/AIDS program. *International Public Management Journal*, 20(1), 160–181.
- Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 22(1), 1–30.
- Hirschman, A. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Harvard University Press.

- Hoffman, E., McCabe, K., & Smith, V. (2008). Chapter 46 reciprocity in ultimatum and dictator games: An introduction. In C. R. Plott & V. L. Smith (Eds.), *Handbook of Experimental Economics Results* (pp. 411-416). Elsevier.
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3–19.
- Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, 67(3), 247–257.
- Hsee, C. K. (1998). Less is better: When low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making*, 11(2), 107–121.
- Hsee, C. K., Blount, S., Loewenstein, G. F., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, 125(5), 576–590.
- Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science* 5(4), 343–355.
- Hvidman, U. (2019). Citizens' evaluations of the public sector: Evidence from two large-scale experiments. *Journal of Public Administration Research and Theory*, 29(2), 255–267.
- Hvidman, U., & Andersen, S. C. (2015). Perceptions of public and private performance: Evidence from a survey experiment. *Public Administration Review*, 76(1), 111-120.
- James, O. (2011a). 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. (2011b). Managing citizens' expectations of public service performance: Evidence from observation and experimentation in local government. *Public Administration*, 89(4), 1419–1435.
- 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.
- Kahneman, D. (2003). Maps of bounded rationality. *American Economic Review*, 93(5), 1449–1475.
- List, J. A. (2002). Preference reversals of a different kind: The “more is less” phenomenon. *American Economic Review*, 92(5), 1636–1643.
- Miller, G. J., & Whitford, A. B. (2007). The principal's moral hazard: Constraints on the use of incentives in hierarchy. *Journal of Public Administration Research & Theory*, 17(2), 213–233.
- Moe, T. M. (1984). The new economics of organization. *American Journal of Political Science*, 28(4), 739–777.
- Moynihan, D. P. (2008). *The dynamics of performance management: Constructing information and reform*. Georgetown University Press.
- Mutz, D. C., & Pemantle, R. (2015). Standards for experimental research: Encouraging a better understanding of experimental methods. *Journal of Experimental Political Science*, 2(2), 192–215.
- Mutz, D. C., Pemantle, R., & Pham, P. (2019). The perils of balance testing in experimental design: Messy analyses of clean data. *American Statistician*, 73(1), 32–42.

- Olsen, A. L. (2017). Compared to what? How social and historical reference points affect citizens' performance evaluations. *Journal of Public Administration Research and Theory*, 27(4), 562–580.
- Ostrom, E. (1973). The need for multiple indicators in measuring the output of public agencies. *Policy Studies Journal*, 2(2), 85–92.
- Rainey, H. G. (2009). *Understanding and managing public organizations*. Jossey-Bass.
- Selden, S. C., & Sowa, J. E. (2004). Testing a multi-dimensional model of organizational performance: Prospects and problems. *Journal of Public Administration Research and Theory*, 14(3), 395–416.
- Sundhedsstyrelsen. (2020). COVID-19. Risikovurdring, strategi og tiltag ved epidemi i Danmark. www.sst.dk
- van Dooren, W., Bouckaert, G., & Halligan, J. (2015). *Performance management in the public sector*. Routledge.
- van Dooren, W., & van de Walle, S. (2008). *Performance information in the public sector: How it is used*. Palgrave Macmillan.
- Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P. G. T., Fu, H., Dighe, A., Griffin, J. T., Baguelin, M., Bhatia, S., Boonyasiri, A., Cori, A., Cucunubá, Z., FitzJohn, R., Gaythorpe, K., ... Ferguson, N. M. (2020). Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, 20(6), 669–677.
- Walker, R. M., Boyne, G. A., & Brewer, G. A. (2010). *Public management and performance: Research directions*. Cambridge University Press.

Walker, R. M., Lee, M. J., James, O., & Ho, S. M. Y. (2018). Analyzing the complexity of performance information use: Experiments with stakeholders to disaggregate dimensions of performance, data sources, and data types. *Public Administration Review*, 78(6), 852–863.

Webeck, S., & Nicholson-Crotty, S. (2020). How historical and social comparisons influence interpretations of performance information. *International Public Management Journal*, 23(6), 798-821.

Accepted author version

Appendices

Appendix A

Table A1. Balance of pre-treatment background variables across treatment groups

	Separate evaluation		Joint evaluation		<i>Difference</i>	P-value
	Mean	N	Mean	N		
Right wing	42.5%	658	41.1%	1,240	1.4	.562
At risk	32.3%	694	32.5%	1,331	-.2	.935
Male	46.8%	694	50.9%	1,331	-4.1	.079
Age	53.6	694	52.5	1,331	1.04	.170
Has college degree	80.2%	683	79.6%	1,311	.6	.711

Notes: Balancing tests do not represent a test of the randomization (Mutz et al., 2019; Mutz & Pemantle, 2015) but can provide an overview of the data.

Table A2. Example of a choice between two strategies (joint evaluation)

	Strategy A	Strategy B
Percentage of infected expected to die from COVID-19	0.5-1%	3-5%
Overall economic cost of COVID-19	More than 2008 financial crisis	Less than 2008 financial crisis
Access to tests	Seriously ill	All in need
Distribution of costs	Equally distributed across groups	Low wage-earners and vulnerable groups hit worst
Access private cell information to reduce the spread of the coronavirus	Yes	No

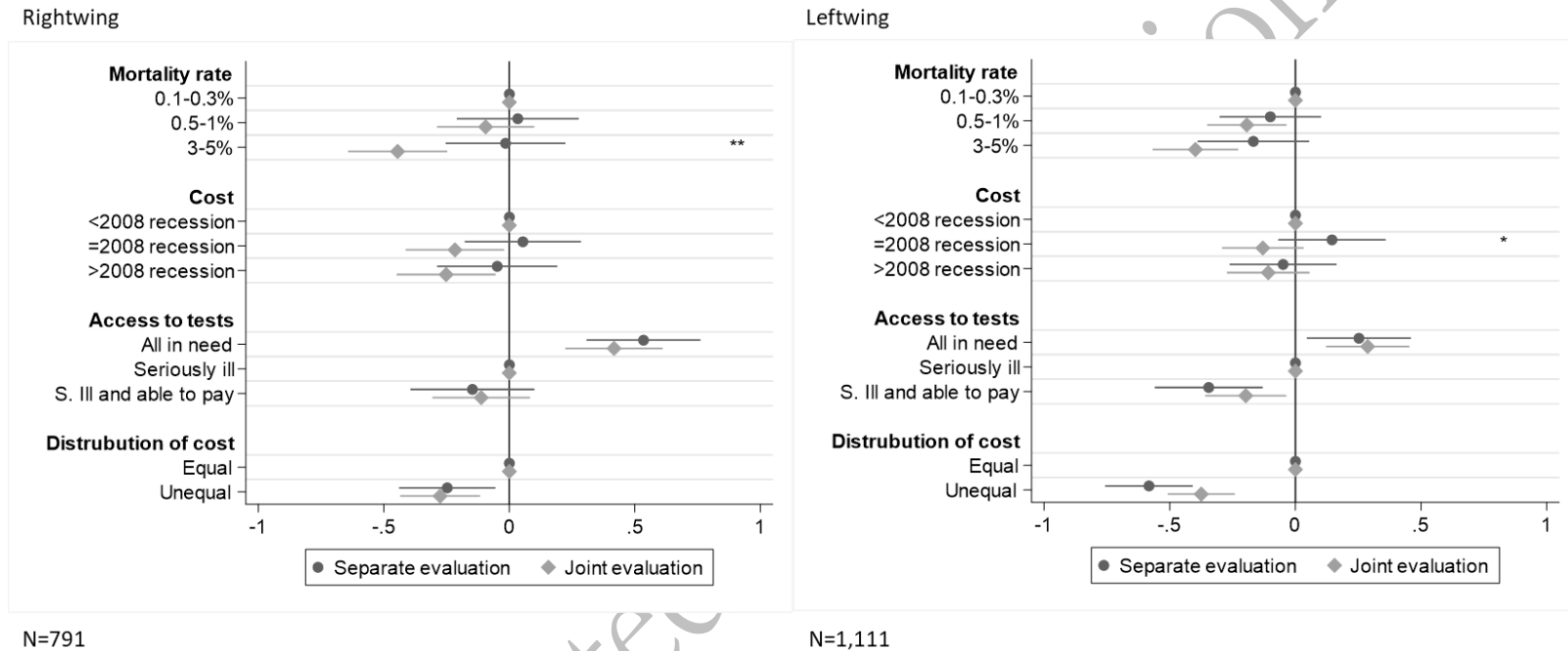
Table A3. Regression tables for separate and joint evaluations. Full results

	Models	
	Separate evaluation	Joint evaluation
<i>Mortality rate</i>		
0.1-0.3%	(reference)	(reference)
0.5-1%	-.0762 (-.94)	-.159* (-2.49)
3-5%	-.109 (-1.29)	-.426*** (-6.30)
<i>Cost</i>		
<2008 recession	(reference)	(reference)
2008 recession	.123 -1.49	-.130* (-1.99)
>2008 recession	-.0135 (-.16)	-.157* (-2.38)
<i>Access to tests</i>		
All in need	.390*** (4.86)	.349*** (5.28)
Seriously ill	(reference)	(reference)
Seriously ill and those able to pay	-.250** (-2.88)	-.174** (-2.70)
<i>Distribution of cost</i>		
Equal	(reference)	(reference)
Unequal	-.433*** (-6.40)	-.329*** (-6.15)
<i>Use of cell data</i>		
No	(reference)	(reference)
Yes	-.253*** (-3.74)	-.234*** (-4.37)
Constant	.605*** (5.88)	.914*** (10.45)
N	694	1,331

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed test). Z-statistics appear in the parentheses.

Appendix B.

Figure B1. The marginal effect of performance information on effectiveness, cost and equality for ratings of COVID-19 strategies. Subgroup analysis. Right-wing and left-wing voters



Notes: Differences between separate and joint evaluations significant at * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Respondents are defined as right wing if they voted for one of the following parties at the last election (Liberal Party (V), Conservative Party (C), Liberal Alliance (I), Danish People's Party (O), and New Right (D)). Respondents are defined as left wing if they voted for one of the following parties at the last election (Social Democratic Party (A), Social Liberal Party (B), Alternative (Å), Socialist People's Party (F), and Red-Green Alliance (Ø)).