Bert George
Bram Verschuere
Ellen Wayenberg
Bishoy Louis Zaki
Ghent University

COVID-19 Viewpoint

A Guide to Benchmarking COVID-19 Performance Data

Bert George is professor of public strategy and behavior in the Department of Public Governance and Management at Ghent University in Belgium. Previously, he held a tenured position in the Department of Public Administration and Sociology at Erasmus University Rotterdam in the Netherlands. His research focuses on strategy, behavior, and performance in public organizations and networks using experimental, observational, and meta-analytical research methods. He is an editorial board member of *Public Administration Review*.
Email: bert.george@ugent.be

Bram Verschuere is professor of public administration and management in the Department of Public Governance and Management at Ghent University in Belgium. His research interests focus on the management of public and civil society organizations, welfare policy, and citizen participation. He is a member of the steering committee of the European Group for Public Administration.
Email: bram.verschuere@ugent.be

Ellen Wayenberg is professor of public policy in the Department of Public Governance and Management at Ghent University in Belgium. She specializes in public policy and administration with a specific interest for policy analysis and evaluation, local government, and multilevel governance. Ellen is one of the co-chairs of the European Group for Public Administration's Study Group on Regional and Local Government and was actively involved in COST Action IS1207: Local Public Sector Reforms: An International Comparison.
Email: ellen.wayenberg@ugent.be

Bishoy Louis Zaki is a PhD candidate in the Department of Public Governance and Management at Ghent University in Belgium. With a primary research focus on policy learning, he has over 12 years of experience in public management, policy, and consulting with governments and public international organizations worldwide. He has also led several local and international capacity building and policy transfer programs.
Email: bishoy.zaki@ugent.be

Abstract: *If the COVID-19 pandemic has taught us anything, it is that policy makers, experts, and public managers need to be capable of interpreting comparative data on their government's performance in a meaningful way. Simultaneously, they are confronted with different data sources (and measurements) on COVID-19 without necessarily having the tools to assess these sources strategically. Because of the speed with which decisions are required and the different data sources, it can be challenging for any policy maker, expert, or public manager to make sense of how COVID-19 has an impact, especially from a comparative perspective. Starting from the question "How can we benchmark COVID-19 performance data across countries?," this article presents important indicators, measurements, and their strengths and weaknesses, and concludes with practical recommendations. These include a focus on measurement equivalence, systems thinking, spatial and temporal thinking, multilevel governance, and multimethod designs.*

While evidence-based policy and management have long been on the agenda of academic theory (e.g., Rousseau 2006; Sanderson 2002) and research (e.g., Hall and Van Ryzin 2019; Jennings and Hall 2012), the recent COVID-19 crisis has definitely put evidence use in decision-making on the agenda of policy makers, experts, and public managers worldwide at all levels of government. Data visualizations demonstrating distributions of COVID-19-related hospital intakes over time in relation to the maximum intensive care unit capacity are widespread under the motto "flatten the curve." Similarly, organizations and sources such as the Organisation for Economic Co-operation and Development (OECD), the World Health Organization (WHO), the University of Oxford, and Johns Hopkins University have been leading the COVID-19 "data revolution" by providing comparative indicators, measurements, and analyses. Policy makers, experts, and public managers have to use these data as part of their crisis management approach to (1) identify best practices from other governments that could be meaningful for them as well and (2) assess how their government is doing and what might explain variation between governments.

Though these initiatives are aimed at ensuring that policy and management decisions concerning COVID-19 are grounded in available evidence, several issues emerge. First, there have been—so far— limited attempts to integrate data sources into one overarching model. This is problematic as it inhibits a thorough holistic understanding of the causes and consequences underlying the COVID-19 crisis. Second, data have been found to sometimes lack reliability and, especially, comparability between governments, making it challenging to compare performance across governments in a meaningful manner even on the same continent. Third, different data sources have popped up, sometimes measuring the same concepts but differently.

This article aims to offer some guidance to policy makers, experts, and public managers at all levels of government navigating the data jungle that emerged with the COVID-19 crisis and hopes to stimulate *strategic thinking* among these actors when benchmarking their own government's performance based on COVID-19 performance data (Bryson and George 2020).

We focus on country-level performance data and seek to answer the question, How can we benchmark COVID-19 performance data across countries? A guide for policy makers, experts, and public managers is developed by following a four-step process: the identification of relevant indicators, the selection of relevant measurements tied to these indicators, the analysis of strengths and weaknesses underlying each measurement, and the elaboration of overarching practical recommendations. The guide acts as a decision-making tool that can help policy makers, experts, and public managers to make sense of their government's performance during the COVID-19 crisis. In doing so, the guide hopes to help them more thoroughly understand the crisis, develop more crisis and risk capacity in the future, and better prepare for, as well as respond to, future crises.

Importantly, while our guide focuses on country-level performance data, we fully realize that this focus is also part of the problem we are addressing here: many comparative data sources are aggregated at the national level, but, depending on the politico-administrative system in a country, it is not always federal or national-level policy makers, experts, and public managers leading the battle against COVID-19. For instance, in the United States, the federal government left many decisions to the states as opposed to playing a strong leadership role itself (Kettl 2020). In other federal countries, such as Belgium and Germany, we observe tensions between the federal government and regional governments, the former aiming at nationwide policies and the latter relying on regional competences to develop their own and sometimes conflicting policies. Also, local governments worldwide are dealing with the consequences of COVID-19, including, for example in the United States, planning their next budget cycle (Maher, Hoang, and Hindery 2020). Similarly, engaging nonstate actors such as private resources and volunteers has proven an effective strategy in Taiwan (Huang 2020).

The effort we propose here in benchmarking performance data to inform decision-making processes is underpinned by mostly national-level data sources. But we believe this can provide a starting point for further analysis at other levels of government. too, if (or when) performance data at these levels are (or become) available. Moreover, we ensure that our defined recommendations are relevant for policy makers, experts, and public managers at all levels of government who seek to benchmark their government's performance based on COVID-19 performance data.

In what follows, the development process is discussed, and the guide—including its indicators, measurements, strengths and weaknesses, and data sources—is presented. Finally, *measurement equivalence*, *systems thinking*, a *temporal* and *spatial* perspective, *multilevel governance*, and *multimethod designs* are discussed as important attention points for further benchmarking initiatives.

Developing a Guide

Two central principles based on findings from recent meta-analyses published in *Public Administration Review* lie at the heart of the developed guide: (1) for performance management to enhance public service performance, it is important that performance data can be (and is) benchmarked between different entities (Gerrish 2016); and (2) decisions grounded in an understanding of capacity and environment, strengths and weaknesses, and systemic, data-based analyses are more likely to enhance public service performance (George, Walker, and Monster 2019). These principles run through the entire development process of the guide to benchmarking COVID-19 performance data.

The starting point of the development process was the research question: How can we benchmark COVID-19 performance data across countries? This question was triggered by discussions among the authors and other colleagues concerning the rather unbalanced and not necessarily valid manner in which COVID-19 performance data were being presented in the media and by policy makers, as well as in the performance rankings developed by research institutions and international organizations.

To answer our research question, the first step was to define and operationalize performance data. Public management looks at

performance as a multidimensional concept, including a range of dimensions, sources, and measurement types (Walker, Boyne, and Brewer 2010). A core theme binding public management studies together is the idea that strategy and policy, capacity, environment, output, and outcome are interrelated in a complex causal chain (Pollitt and Bouckaert 2017). Typically, strategy and policy are implemented with the aim of enhancing output and, eventually, outcome, but whether or not this enhancement occurs is contingent on capacity and environment (O'Toole and Meier 2015). It is thus not enough for policy makers, experts, and public managers to focus only on output and outcome. For a better systemic—or causes-consequences—understanding of COVID-19, performance indicators related to strategy and policy, capacity, environment, output, and outcome should be looked at simultaneously to come up with a more holistic picture (which is all the more crucial given the complex and multidimensional nature of the problem at hand).

The second step was finding relevant measurements of strategy and policy, capacity, environment, output, and outcome. The choice was made to focus only on data sources coming from "official" research institutions or international organizations such as the WHO and the OECD. A variety of measurements were considered and discussed among the authors until those deemed most relevant for the issue at hand were retained.
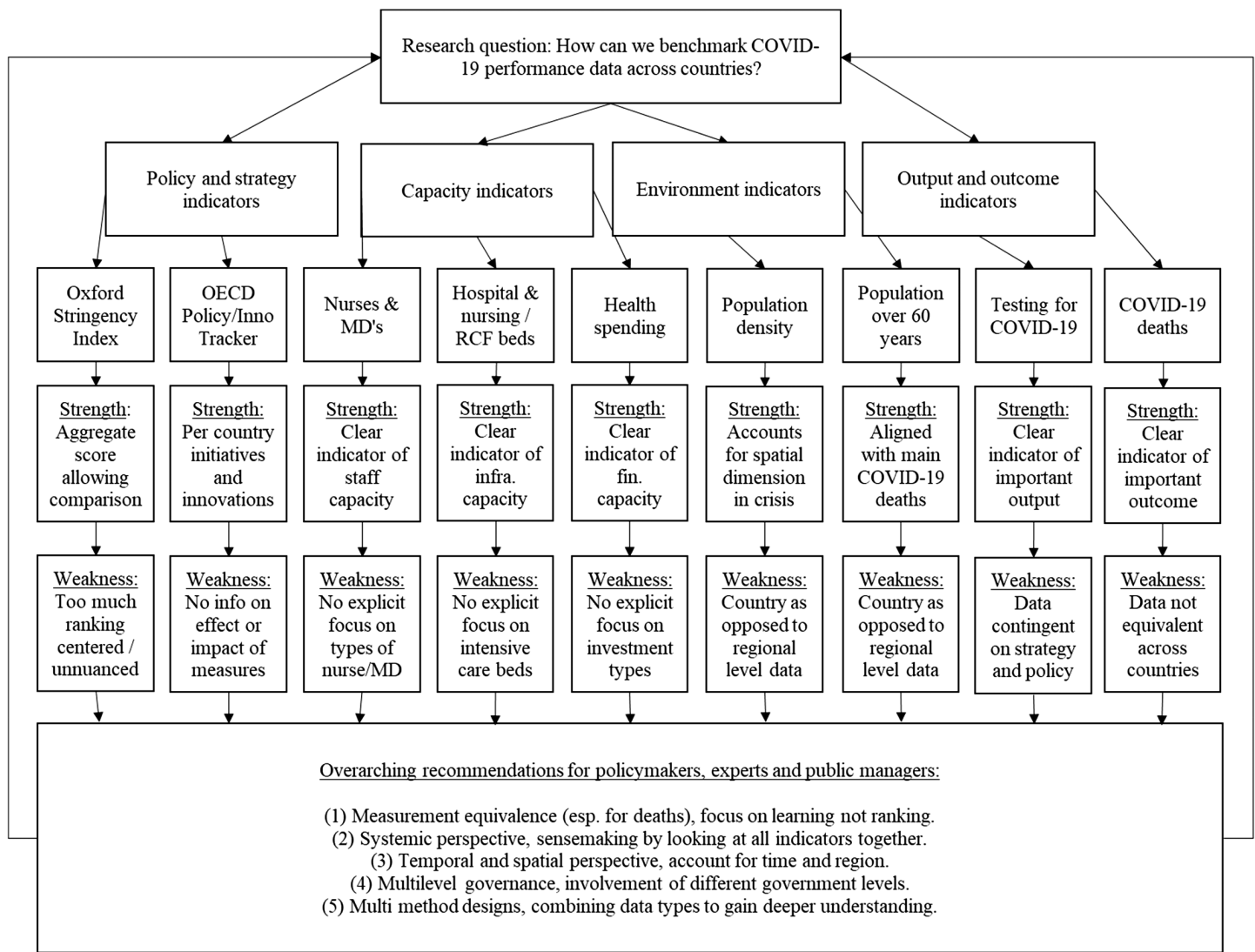
The third step was to analyze each selected measurement and identify its strengths and weaknesses for policy makers, experts, and public managers who might use these tools in the near future or already are using them. Finally, based on an overarching analysis of all the identified strengths and weaknesses, the entire process of identifying indicators and measurements, and a recent elaboration of what constitutes *strategic thinking* in public administration (Bryson and George 2020), key recommendations for future benchmarking initiatives were distilled. Figure 1 offers a visual representation of the developed guide.

Figure 1 shows that several measurements exist to assess strategy and policy, capacity, environment, output, and outcome indicators.

Strategy and Policy

The measurement of strategy and policy includes both a quantitative and a qualitative component. The Oxford COVID-19 Government Response Tracker (Blavatnik School of Government, University of Oxford) offers a per-country score based on 17 indicators of responses from governments to COVID-19. Eight indicators center on containment and closure policy actions undertaken by governments (e.g., school closures, travel bans), whereas the other indicators center on economic efforts (e.g., fiscal or monetary measures) or health system policies (e.g., testing regimes) (Hale et al. 2020). Because of its quantitative and aggregative nature, it is an ideal tool to identify top-, average-, and poor-performing countries and to assess the underlying best practices. However, this is not without its downsides, as a ranking of countries is never far away based on the index, and thus it can become more of a political tool used for credit taking and blame avoidance strategies as opposed to really stimulating learning and knowledge sharing (Nielsen and Baekgaard 2015).

The OECD is an international organization that focuses on building better policies for better lives; it includes 37 member

Figure 1 includes a flowchart with the following text:

Research question: How can we benchmark COVID-19 performance data across countries?

- Policy and strategy indicators
- Capacity indicators
- Environment indicators
- Output and outcome indicators

Indicator	Strength	Weakness
Oxford Stringency Index	Aggregate score allowing comparison	Too much ranking centered / unnuanced
OECD Policy/Inno Tracker	Per country initiatives and innovations	No info on effect or impact of measures
Nurses & MD's	Clear indicator of staff capacity	No explicit focus on types of nurse/MD
Hospital & nursing / RCF beds	Clear indicator of infra. capacity	No explicit focus on intensive care beds
Health spending	Clear indicator of fin. capacity	No explicit focus on investment types
Population density	Accounts for spatial dimension in crisis	Country as opposed to regional level data
Population over 60 years	Aligned with main COVID-19 deaths	Country as opposed to regional level data
Testing for COVID-19	Clear indicator of important output	Data contingent on strategy and policy
COVID-19 deaths	Clear indicator of important outcome	Data not equivalent across countries

Overarching recommendations for policymakers, experts and public managers:

(1) Measurement equivalence (esp. for deaths), focus on learning not ranking.
(2) Systemic perspective, sensemaking by looking at all indicators together.
(3) Temporal and spatial perspective, account for time and region.
(4) Multilevel governance, involvement of different government levels.
(5) Multi method designs, combining data types to gain deeper understanding.

OECD = Organization for Economic Cooperation and Development.

Figure 1   A Guide to Benchmarking COVID-19 Performance Data OECD = Organisation for Economic Co-operation and Development.

countries (including the United States). It offers two tools focused on strategies and policies across countries. The OECD Policy Tracker and Innovation Tracker are more qualitative in nature and offer a per-country overview of implemented policies as well as specific innovations used by countries to tackle COVID-19 related issues. Policy makers, experts, and public managers can use these to identify best practices in countries and how these were realized. For instance, the Policy Tracker shows whether and when countries initiated quarantine or confinement initiatives and closed schools and universities, and the Innovation Tracker allows the identification of innovations (e.g., the initiation of a public-private partnership) at different levels of government (national or federal, local, regional, etc.) that have helped tackle specific consequences of the COVID-19 pandemic. However, the trackers lack a concrete assessment of the impact or effect achieved through the described innovations and policy initiatives, and these always need to be contextualized to ensure applicability and relevance in one's own context (e.g., differences between politico-administrative regimes; see Kuhlmann and Wollmann 2019; Pollitt and Bouckaert 2017).

Capacity and Environment
The capacity and environment measurements are quantitative in nature and were all derived from databases provided by the WHO. The capacity measurements have to be tied to the health care system in place in each country, and thus they include staff numbers, especially of health care professionals in direct contact with patients (e.g., nurses and medical doctors), the number of beds available in hospitals as well as residential care facilities, and health care spending. All of these measurements offer insights into the staff, infrastructural, and financial capacity underlying the health care system of a country. However, these do not necessarily offer insights into the types of staff (e.g., specialists in respiratory systems), beds (e.g., intensive care), and investments (e.g., ventilators) that are particularly relevant for COVID-19.

The environment measurements center on two measurements that are particularly relevant for the COVID-19 case, namely, the share of the population over age 60 (i.e., the biggest at-risk

group) and the population density—which is especially relevant for the diffusion rate of COVID-19 infections. While these data help identify the vulnerability of a country, one limitation is that data often are not presented at a more fine-grained regional level; rather, broader country averages are offered. Or one needs to look at databases offered by each country specifically to assess regional level data, which in practice is quite a burdensome activity that might be hampered by language, transparency, and technological barriers.

Output and Outcome

Finally, the output and outcome measurements are quantitative in nature. A clear output indicator is the number of tests for COVID-19 being conducted by a country. Again, this is a measurement that is clearly relevant to the context; however, it should always be looked at together with the other measurements (especially concerning strategy and policy), as the number of tests is highly contingent on the strategies and policies employed by a country. Moreover, it is self-reported data and, as such, might be prone to measurement error derived from the different ways in which tests are conducted across countries.

As an outcome measure, COVID-19-related deaths is by far the most reported measurement both in the media as well as by policy makers. Clearly, minimizing the number of deaths is crucial thus making this measurement very relevant. At the same time, the COVID-19 death rate is self-reported by countries, and there are many different ways in which deaths are counted. This hampers the measurement equivalence underlying this outcome measurement, and policy makers, experts, and public managers should first ensure that they are indeed comparing themselves with countries that use the same (or at least a very similar) way of measuring COVID-19 deaths. Table 1 offers an overview of the sources used for each measurement.

Table 1 Performance Data Sources on COVID-19

Indicator	Measurement	Source*
Strategy and policy	Oxford Stringency Index	https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker
	OECD Policy Tracker	http://www.oecd.org/coronavirus/en/
	OECD Innovation Tracker	https://oecd-opsi.org/covid-response/
Capacity	Nurses and MDs	https://apps.who.int/gho/data/node.home
	Hospital and nursing/residential care facility beds	https://apps.who.int/gho/data/node.home
	Health spending	https://apps.who.int/gho/data/node.home
Environment	Population density	https://apps.who.int/gho/data/node.home
	Population over age 60	https://apps.who.int/gho/data/node.home
Output and outcome	Testing for COVID-19	https://ourworldindata.org/covid-testing
	COVID-19 deaths	https://coronavirus.jhu.edu/map.html

OECD = Organisation for Economic Co-operation and Development.
*All online data sources were consulted on April 28, 2020.

Measurement Equivalence, Systems Thinking, Time and Space, Multilevel Governance, and Multimethod Designs

Although the guide we present aims to enable meaningful benchmarking initiatives between governments concerning COVID-19, it needs to be handled with care. To ensure that data are actually used to enhance learning and knowledge sharing, the proposed guide suggests the following recommendations aimed at policy makers, experts, and public managers at all levels of government, drawing on our initial initiative of benchmarking country-level performance data:

- Be careful when benchmarking your own government's performance. Take into account (a lack of) *measurement equivalence* between data sources—especially for COVID-19 deaths performance data. Moreover, use these data not for ranking purposes but rather for learning and knowledge sharing. It is particularly important to be wary of government performance rankings published by international organizations and research institutions, as the data used are not always comparable and can result in rather narrow comparisons as opposed to a deep understanding of different performance dimensions.
- Always interpret COVID-19 performance data within a broader *systemic perspective* as opposed to looking at indicators separate from each other. Policy and strategy, capacity, environment, output, and outcome together help make sense of per-government performance and between-governments variance in that performance. In other words, assess these data using a series of cause-effect relationships to better unravel the broader system underlying the data.
- Acknowledge the *temporal* and *spatial* perspective underlying COVID-19 performance data. This implies attention to how indicators evolve over time (and why), differences in performance that might be the result of regional characteristics as opposed to typical country boundaries (and why), and using the most recent data that are (or soon become) available.
- Acknowledge the *multilevel governance* system in addressing policy challenges like the COVID-19 pandemic. Governments at all levels—local, state, regional, federal, national, and even supranational—as well as nonstate actors such as nonprofit organizations and civil society are involved. Depending on the type of politico-administrative regime, the nature of intergovernmental policy making and collaboration, and the nature of government–civil society collaboration can be very different. Aggregate, country-level performance data do not necessarily account for this multilevel and collaborative governance reality. When feasible, these data should be complemented with regional, state, or local level data to identify within-country variance in performance at different levels of government (and understand why).
- Do not only rely on quantitative performance data but adopt *multimethod designs*. As is clear from the Innovation Tracker provided by the OECD, there are many initiatives ongoing worldwide at all levels of government to battle the COVID-19 pandemic. Such initiatives are not captivated by quantitative data but are reported as case studies or best practices. While the quantitative data can help identify differences in performance between governments, more qualitative data

can help understand exactly why and how these differences emerge, and which lessons can be learned for other contexts. We can think of, for example, a qualitative assessment of how the nature of multilevel governance determines policy decision-making (e.g., Bouckaert et al. 2020), and eventually performance.

While these recommendations can help policy makers, experts, and public managers benchmark their own government with others using COVID-19 performance data, some limitations of this article have to be acknowledged. First, measurements tied to the overall indicators were selected based on discussions and learning processes between the authors. Quite obviously, many other measurements exist and can be useful for future reference. Policy makers, experts, and public managers should thus be encouraged to not consider the selected measurements as an exhaustive list but, rather, a starting point for further analysis.

Second, the COVID-19 performance data are highly dynamic and any reader of this article should check whether the identified sources are still up-to-date and/or whether new and perhaps better sources have emerged. One example is the recent initiative from, among others, the *Financial Times*[1] and EuroMOMO[2] to identify excess mortality linked to COVID-19, which could be a more robust and comparable indicator of COVID-19 outcomes than absolute number of deaths.

Finally, this guide can act as a strategy tool that enables benchmarking initiatives, but it does not replace the cognitive capacity underlying the people actually making the decisions, conducting the analyses, and debating them. Several studies have shown that heuristics, psychological characteristics, group composition, and group dynamics influence the quality of decision-making in government (e.g., Battaglio et al. 2019; George 2020), and the developed guide is unlikely to yield benefits in decision-making teams where any of the aforementioned cognitive and social aspects are neglected.

Notes

1   John Burn-Murdoch, Valentina Romei, and Chris Giles, "Global Coronavirus Death Toll Could Be 60% Higher than Reported," *Financial Times*, April 26, 2020, https://www.ft.com/content/6bd88b7d-3386-4543-b2e9-0d5c6fac846c (accessed April 29, 2020).
2   EuroMOMO, https://www.euromomo.eu/ (accessed April 30, 2020).

References

Battaglio, R. Paul, Jr., Paolo Belardinelli, Nicola Bellé, and Paolo Cantarelli. 2019. Behavioral Public Administration *ad fontes*: A Synthesis of Research on Bounded Rationality, Cognitive Biases, and Nudging in Public Organizations. *Public Administration Review* 79(3): 304–20.

Bouckaert, Geert, Steven van Hecke, Davide Galli, Sabine Kuhlmann, and Renate Reiter. 2020. European Coronationalism? A Hot Spot Governing a Pandemic Crisis. *Public Administration Review*. Published online May 22. https://doi.org/10.1111/puar.13242.

Bryson, John M., and Bert George. 2020. Strategic Management in Public Administration. In *Oxford Encyclopedia of Public Administration*. https://doi.org/10.1093/acrefore/9780190228637.013.1396.

Gerrish, Ed. 2016. The Impact of Performance Management on Performance in Public Organizations: A Meta-Analysis. *Public Administration Review* 76(1): 48–66.

George, Bert. 2020. Successful Strategic Plan Implementation in Public Organizations: Connecting People, Process, and Plan (3Ps). *Public Administration Review*. Published online April 13. https://doi.org/10.1111/puar.13187.

George, Bert, Richard M. Walker, and Joost Monster. 2019. Does Strategic Planning Improve Organizational Performance? A Meta-Analysis. *Public Administration Review* 79(6): 810–9.

Hale, Thomas, Noam Angrist, Beatriz Kira, Anna Petherick, Toby Phillips, and Samuel Webster. 2020. Variation in Government Responses to COVID-19. Working Paper, Blavatnik School of Government, University of Oxford. https://www.bsg.ox.ac.uk/research/publications/variation-government-responses-covid-19 [accessed June 7, 2020].

Hall, Jeremy L., and Gregg G. Van Ryzin. 2019. A Norm of Evidence and Research in Decision-Making (NERD): Scale Development, Reliability, and Validity. *Public Administration Review* 79(3): 321–9.

Huang, Irving Yi-Feng. 2020. Fighting COVID-19 through Government Initiatives and Collaborative Governance: The Taiwan Experience. *Public Administration Review*.

Jennings, Edward T., and Jeremy L. Hall. 2012. Evidence-Based Practice and the Use of Information in State Agency Decision Making. *Journal of Public Administration Research and Theory* 22(2): 245–66.

Kettl, Donald F. 2020. States Divided: The Implications of American Federalism for COVID-19. *Public Administration Review*.

Kuhlmann, Sabine, and Hellmut Wollmann. 2019. *Introduction to Comparative Public Administration*. Northampton: Edward Elgar.

Maher, Craig S., Trang Hoang, and Anne Hindery. 2020. Fiscal Responses to COVID-19: Evidence from Local Governments and Nonprofits. *Public Administration Review*.

Nielsen, Poul A., and Martin Baekgaard. 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): 545–69.

O'Toole, Laurence J., Jr., and Kenneth J. Meier. 2015. Public Management, Context, and Performance: In Quest of a More General Theory. *Journal of Public Administration Research and Theory* 25(1): 237–56.

Pollitt, Christopher, and Geert Bouckaert. 2017. *Public Management Reform: A Comparative Analysis-into the Age of Austerity*. Oxford: Oxford University Press.

Rousseau, Denise M. 2006. Is There Such a Thing as "Evidence-Based Management"? *Academy of Management Review* 31(2): 256–69.

Sanderson, Ian. 2002. Evaluation, Policy Learning and Evidence-Based Policy Making. *Public Administration* 80(1): 1–22.

Walker, Richard M., George A. Boyne, and Gene A. Brewer. 2010. *Public Management and Performance: Research Directions*. Cambridge: Cambridge University Press.