tpa

Identifying and interpreting government successes: An assessment tool for classroom use

Scott Douglas ⓘ, Paul 't Hart ⓘ and Judith Van Erp
Utrecht School of Governance, Utrecht University, Netherlands

Abstract
Journalists, politicians, watchdog institutions, and public administration scholars devote considerable energy to identifying and dissecting failures in government. Studies and case-studies of policy, organizational, and institutional failures in the public sector figure prominently in public administration curriculums and classrooms. Such a focus on failures provides students with cautionary tales and theoretical tools for understanding how things can go badly wrong. However, students are provided with less insights and tools when it comes to identifying and understanding instances of success. To address this imbalance, this article offers students a framework to systematically identify, comprehensively assess and carefully interpret instances of successful public governance. The three-stage design of the funnel introduces students to relevant debates and literatures about meaningful public outcomes, the prudent use of public power, and the ability to sustain performance over time. The articles also discuss how this framework can be used effectively in classroom settings, helping teachers to stimulate reflection on the key challenges of assessing and learning from successes.

Corresponding author:
Scott Douglas, Utrecht School of Governance, Utrecht University, Bijlhouwerstraat 6, Utrecht 3508 TC, Netherlands.
Email: s.c.douglas@uu.nl

Introducing students of public administration to the systematic study of success

There is an abundance of analytical frameworks and tools to learn from governance "disasters" (Gray and 't Hart, 1998; Hall, 1981), blunders (Jennings et al., 2018; King and Crewe, 2013), policy failures (Bovens and 't Hart, 1996; Light, 2014; Opperman and Spencer, 2016), blind spots (Bach and Wegrich, 2018), and blame games (Hinterleitner, 2018; Hood, 2010). Studying these "dark sides" can teach us what to avoid, prevent and contain when crafting public institutions and designing and implementing public policies. Yet, while there is need for "learning what to avoid" from the study of government failures and crises (McDonald, 2021), there is also a need for "learning what to aspire to and emulate" by focusing on government successes.

Our current academic practices about government and public administration are not equally attuned to spotting and naming successes as they are to finding faults and blaming public officials and agencies for them (Luetjens and 't Hart, 2019).This in contrast to the field of business management where the language of success is paramount and carefully analyzed cases of successful businesses permeate both research and teaching. If we want to equip students of public administration to study successes just as well as we prepare them for studying failure, we need a conceptual apparatus to help them systematically identify, comprehensively assess, and carefully interpret instances of successful governance.

In this article we contribute to the emerging "positive turn" in public administration scholarship (Compton et al., 2022; Douglas et al., 2021) by presenting a framework for assessing, analyzing, and interpreting cases of successful governance that can be used in classroom settings. The framework outlines different dimensions of success along which (sets of) cases can be identified, assessed, analyzed, and learned from. It provides clear normative standards that enable systematic classroom discussion about what we value in government, what standards we apply to assess public organizations and public programs, and what methodological and analytical repertoires we might use in efforts to learn from positives.

The article first offers students a framework to systematically identify, comprehensively assess, and carefully interpret instances of successful public governance. The three-stage design of the funnel introduces students to relevant debates and literatures about meaningful public outcomes, the prudent use of public power, and the ability to sustain performance over time. The articles then also offer teachers an insight in how this framework can be used effectively in classroom settings. Based on our first applications of the model in public administration courses, we discuss how this model and its application to cases of governance can stimulate reflection on the key challenges of assessing and learning from successes. The following sections first briefly outline the genesis of the framework, then discuss the content of the framework, and finally provide a step by step approach for its application in the classroom.

Designing and using a framework for identifying and assessing government success

The framework was developed through both a use of the existing literature on government and public governance and the practical application of the funnel in the classroom. Theoretically, the funnels seeks to both build upon and transcend existing foundational work on positive policy evaluation (Fetterman 2006; Fetterman and Wandersman, 2005; Nielsen et al., 2015); policy success (Compton and 't Hart, 2019; Luetjens et al., 2019; McConnell, 2010); regulatory excellence (Coglianese, 2016); public value creation (Alford et al., 2017; Bryson et al., 2015; Moore, 1995, 2013); and successful collaborative governance and network management (Cristofoli et al., 2017; Dickinson and Sullivan, 2014; Page et al., 2015). Similarly, we integrate research on successful political and public innovation (Hartley et al., 2013; Sørensen, 2017); high-performing and highly reputed public sector organizations (Carpenter, 2001; De Waal, 2010; Goodsell, 2011); exemplary public administrators (Cooper and Wright, 1992); resilient systems (Comfort et al., 2010; Walker and Salt, 2006); and high-reliability systems performing public tasks in high-risk operating environments (Rochlin, 1996; Roe and Schulman, 2008; Weick and Sutcliffe, 2011).

This previous research has generated valuable insights into "what works" across different domains and forms of public governance, but scholars vary widely in how they choose to conceptualize and assess what is desirable about them. The first contribution of the assessment tool is to capture and connect these various conceptualizations of success, by designing a multicriteria funnel for evaluating instances of governance. The second contribution of the framework is to enable a systematic reflective discussion among practitioners and students about success in the public sector and what general lessons, if any, can be drawn from specific cases. On the whole, the framework offers a first step for a more systematic, comprehensive yet nuanced analysis of governance success.

We have integrated the use of the framework in a number of undergraduate, post-graduate, and executive courses on successful public governance. These experiences led to both small adjustment of the funnel, specifically clarifying what we mean with the different dimensions and tests involved, and the development of a practical teachings approach to using this funnel in the classroom. This latter practical application centers using the funnel to provide students with an overview of the different ideas about "success" in publication administration, offering students a tool for finding and assessing cases themselves, the use of additional reports and materials to substantiate their judgments, and ongoing classroom discussion and reflection about how government success can be identified, understood, and promoted. We first discuss the substantive elements of the framework below, before highlighting the practical application.

Understanding the funnel framework

Our ambition is to offer a discursive tool for assessing the nature and degree of government successes. We are specifically interested in "cases of public governance," which we define as discrete bundles of activity aimed at addressing public issues. These

activities can be undertaken solely by government, but also in collaborations between government and other private or community actors. Governance can take the shape of policies, projects, programs, service delivery mechanisms, organizations, and collaborations.

We argue that governance success cannot simply and solely be measured along a limited number of straightforward performance indicators. Governing is a complex social and inherently political activity that itself is shaped by laws and regulations, social norms and expectations, as well as the relative power of different organized interests. Its success or otherwise is always assessed and debated from multiple vantage points and value sets (Bovens et al., 2001). Against this backdrop we surmise that any instance of public governance can be considered completely successful when it fully satisfies three tests: (1) the instance of governance delivers a meaningful public contribution of valuable and valued societal results; (2) the instance of governance demonstrates a prudent and legitimate use of public power; (3) the instance of governance secures a sustained performance over time.

These three tests are not to be thought of as a multi-perspectivist framework of different constitutional powers (law, politics, and management, Rosenbloom, 1983), concurrent or competing values (Hood, 1991; Lindquist and Marcy, 2016) nor as a normative hierarchy (Fischer, 1995). As we will elaborate below, these tests are purposefully constructed as a *three-stage funnel* with each stage comprising a set of evaluative checks. Only cases that pass all three stages of the assessment funnel process with flying colors can be considered to be "complete successes," whereas cases that do not even make it past the first sets of checks can be considered a failure.

As observed by McConnell (2010), when the framework is applied to a randomly selected universe of cases, there will be many "in-between" cases of partial (maximum scores on some but not all tests) or conflicted successes (persistent disagreement among different evaluators about what scores should apply on certain criteria). The funnel framework can thus be a useful evaluation tool for high, low, and medium performing cases. However, our primary interest in offering this tool is its potential for helping student to systematically identify and analyze cases on the high end of the success spectrum.

Importantly, we see this framework as opening up the discussion of success, purposefully offering a rich and eclectic mix of issues and questions. This funnel framework may so provide a useful start for identifying cases of success, but researchers seeking more precise and methodologically pure assessment of their case samples may need to then move on to more rigid assessments frameworks.

Figure 1 presents a visual representation of the funnel framework. We now introduce each of the meta-criteria, the "tests" subsumed under them, and operationalize these into assessment questions which scholars, students, and practitioners can use to assess cases of public governance.
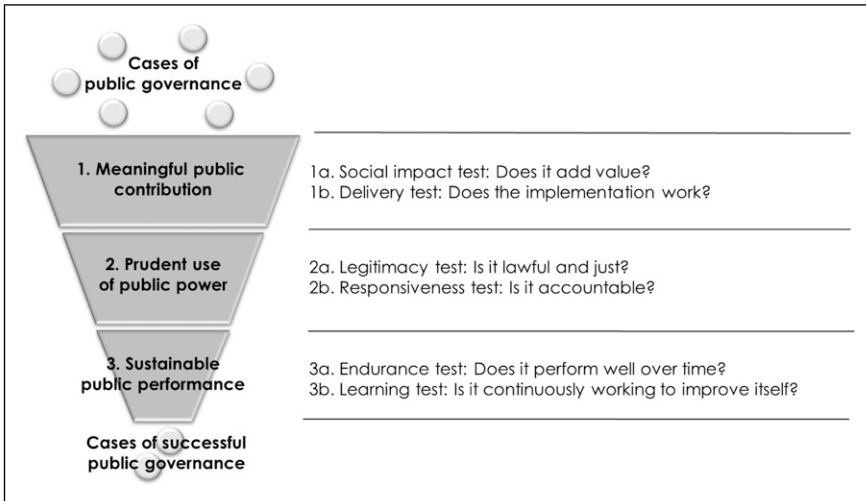
Figure 1. The successful public governance assessment funnel.

Filter 1 - Meaningful public contribution

The social impact test: Does it add value?

Does the pattern of (intended and unintended) social benefits that results from the initiative outweigh its costs? Is this pattern of benefits and costs positively valued across the spectrum of stakeholders?

The first step is to establish whether a particular case of governance adds value to society. This added value can be of a material and non-material nature. Government can contribute to society by providing more housing, faster transport links, better medical care or cleaner environments to citizens, but also by strengthening a sense of community, fostering well-being, or social cohesion. These contributions often come at a cost. Again not only in material terms such as higher taxes, but also in non-material terms such as the expansion of government power and limitation of individual liberties. Moreover, both the benefits and costs of the government action will be a mix of intended and unintended effects, with government initiatives sometimes generating unexpected benefits but also incurring unplanned downsides (Boudon, 2016). For example, the construction of a new highway bypass around a city center has both material and immaterial impacts (time saved by traffic, impact on air quality, costs of construction) and knock-on effects (reducing visitors to the city center shops, neighborhoods cut off city by highway).

The ambition is that the overall benefits of government action outweigh the costs, expanding the aggregate social welfare function, that is, more benefits and fewer costs for society as a whole (Boardman et al., 2017). In specific cases, such expansions can be achieved through innovations or breakthroughs that benefit everyone in society (i.e., Pareto improvement), but often government action will involve a redistribution of benefits

from one group of actors to another group of actors or an infringement of the rights or interests of specific people (Pareto shifts). For example, the development of the vaccine against polio could be seen as "a net gain" for humanity, although debates may swiftly follow about whether everyone can be obliged to receive this vaccination.

The assessment of the value of government action is therefore also a question about the actors that are affected by the policies or initiatives (Moore, 1995). General and abstract cost-benefit analyses may struggle to capture the experiences of the various actors affected by government, and such analyses therefore needs to be complemented with a more contextualized, stakeholder-oriented approach. Such as an exercise reveals to what extent the outcome is considered worthwhile by the stakeholders affected while also gaging the level of satisfaction among diverse groups. Importantly, the different scores cannot be simply averaged out in order to ascertain a net positive contribution. The gain of the many does not necessarily outweigh the pain of the few. This assessment of the distribution of value and costs across stakeholders ultimately requires a normative judgment.

The Delivery test: Does the implementation work?

Are implementation mechanisms and delivery practices evidence-based and appropriately tailored to the context in which the activities take place?

The delivery test assesses whether the implementation and execution of the government intents are carried out in the best possible way. The chosen implementation mechanisms need to be informed by solid evidence and in line with the latest scientific insights (Nutley et al., 2007). For example, a program fighting substance abuse should be informed by the latest insights in what interventions work best to reduce addiction (Miller et al., 2006).

Moreover, the organization of how these interventions are delivered needs to align with the context. Research on policy instruments offers policy designers evidence-based insights for calibrating the settings and indeed the mix of instruments they use to influence the attitudes and behaviors of different target populations (Howlett et al., 2015) and the task. For some public services, market-based implementation through self-regulation, privatization and performance-based management may be a suitable mechanism, but when a well-functioning market and discriminating client base is absent, this may not be the most effective approach. For example, the World Health Organization recommends that each national vaccination program finds the appropriate partner for each geographical and societal context in which it works (World Health Organization, 2017). Health clinics may be the appropriate delivery partners in urban areas, but religious organizations might have the better networks to reach more remote corners of the world.

Filter 2 - Prudent use of public power

The legitimacy test: Is it lawful and just?

Are governance processes and outcomes accordance with the Rule of Law and perceived as just and fair by all stakeholders?

All forms of public governance should be in accordance with the Rule of Law: they should comply with constitutional, international, and domestic law. The Rule of Law protects citizens from abuse of power through arbitrariness and willfulness by restricting discretion of government officials and requiring due process, and thus enhances certainty, predictability, and security between citizens and the government, and among citizens (Tamanaha, 2007). It thus generates trust and provides the soil in which "successful" governance can flourish (Rothstein, 2012). Although opinions differ about the elements that constitute the Rule of Law, it is generally agreed that law must be set forth in advance (be prospective), be public, be general rather than particularistic, be clear, be stable and certain, and be applied to everyone irrespective of person, position, or status. Rule of Law demands are "thicker" than the criterion of legality. Hence, authoritarian rulers may establish policies that are legal in the sense that they are in accordance with national laws, but the Rule of Law demands more than legality. Public governance cannot be considered "successful" when it does not respect substantial values, such as human rights, justice, sustainability, and social equity (Waldron, 2016).

The Rule of Law poses significant challenges in the context of modern governance; complexity; and volatility. Open norms, privatization, decentralization, and collaborative governance are often sought to increase the effectiveness and adaptiveness of policies in complex and dynamic environments but may compromise the Rule of Law requirements of publicness, generality, and stability. Successful governance in such settings requires innovative efforts to communicate laws and policies to make them accessible and intelligible for everyone, including tailor-made outreach to specific target groups and segments of society. Similarly, contrary to the demands of predictability and stability, a certain degree of discretion is necessary in unanticipated situations and changed circumstances, as highly "juridified" systems have strong disadvantages. If "every functional polity must accord some degree of trust and discretion to government officials" (Tamanaha, 2007, 11), this presumes a high degree of trust in the democratic and legal system—one that stretches beyond particular cases of governance.

Moreover, substantive values, such as "justice" and "fairness" are necessarily subjective and context-dependent. For example, a social housing policy safeguards the availability of affordable housing for lower income groups. But whether it should give asylum seekers priority over local residents is a hotly debated issue in western European countries with tight housing markets. Policy making also requires weighing values against each other—investments in sustainable housing often lead to increased rents of social housing and lower affordability. Applying the legitimacy test therefore requires a normative judgment. We therefore add the criterion of legitimacy, which reflects normative acceptance of governance processes as manifested in broad public support for and trust in governance actors appropriate within social norms, values, beliefs, and definitions (Page et al., 2015; Suchman, 1995). Legitimacy does not just emerge but instead is actively crafted and developed in processes of legitimation (Van Assche et al., 2011). Multiple "publics" with heterogeneous and potentially conflicting beliefs, values, and interests exist to be convinced of the legitimacy of a governance process (Prebble, 2018).

The standard for successful legitimization should therefore not be the presence of a complete consensus, but constructive engagement of all stakeholders with the policy.

Yang (2016) characterizes the necessary steps as participation of various groups in the process of articulating interests; legitimation of the informal outcome by translation to formal policies in order to ensure political commitment, and implementation to ensure actual results. The Rule of Law and legitimacy are closely intertwined: There is a mountain of research to show that procedural justice can contribute to perceived legitimacy of outcomes—even in conflict of interest situations (Maguire, 2018; Tyler, 2001).

Rule of Law, legitimacy, justice, and human rights are universal values. Yet, their assessment may pose challenges in developing countries as it may be more difficult to assess "legitimacy" in authoritarian regimes where public discourse is less free or where tensions exist between legality and the Rule of Law. Hungarian laws to curtail judicial independence and freedom of expression, for example, were declared in conflict with the Rule of Law by the European Parliament.[1] The successful governance assessment funnel and its tests have been developed from a western democratic perspective, and may require adjustment when applied in developing countries. We will return to this issue in the reflection section at the end of this paper.

The responsiveness test: Is it accountable?

Do the key public actors involved in a governance practice engage in proactive and responsive account-giving to multiple audiences that allow these to be well-informed about and able to evaluate its merits and progress?

Governance should not only reflect the Rule of Law, but also respond and account to the public by reporting, explaining, and justifying their acts to allow the public to evaluate the success of governance (Behn, 2001). In representative democracies, governments need to be accountable to political principals and responsive to the will of the public as represented by elected politicians (Mulgan, 2014). They also should hold themselves accountable to accountability institutions such as Ombudsmen, Courts of Audit, international bodies, and professional norms and standards setting bodies. Such checks and balances prevent the arbitrary exercise of power. Accountability is central especially in the context of networked relationships, where various parties interact through a variety of competitive, cooperative, negotiated, and command and control arrangements. As state authority is increasingly shared with others, challenges arise for accountability, as formal democratic control mechanisms may not be able to capture network structures and processes.

Hierarchical and professional forms of accountability have long dominated accountability practices inside government. However, such managerial, performance-oriented approaches typically do not assess governance practices against more comprehensive definitions of the public interest (Rosenbloom, 1983). In addition to state-based accountability forums, non-state institutions such as the media, private regulatory actors such as Forest Stewardship Council for sustainable timber; citizen rights NGOs such as Amnesty International or environmental NGOs, hold governments to account in the media or in court. An example of the latter is the Urgenda movement in the Netherlands which successfully litigated against the Dutch State for failing to implement

the goals of the Paris Climate Agreement, effectively forcing it to adopt a more ambitious climate policy.

Although policies and government actors do not need to be substantially responsive to all these audiences, successful governance entails that they be procedurally responsive by being transparent, providing performance information, engaging in dialog, in other words, by accepting responsibility. We surmise that public actors and initiatives that go beyond their formal legal accountability to political principals and develop both proactive and responsive account-giving practices in relation to multiple audiences are more likely to be successful. These account-giving practices are likely to contribute to the better operation of checks and balances and thus a mature and balanced scrutiny of the extent to which the public interest is being served. It is this maturity of information provision, debate, and assessment that increases the public's trust in what is being undertaken on its behalf and thus enhances the reputation of the initiatives and actors involved.

Filter 3 - Sustainable public performance

The Robustness test: Does it perform well over time?

Are considerations of long-term viability given due attention in the institutional design and management of the initiative?

The final step of the assessment funnel focuses on the temporal dimension of good practices and high performance: are they designed to endure, and do they? The pre-occupation is to be able to keep going when others are thrown off course by changing operating and political environments yet while preserving their commitment to the core values and principles that lie at the heart of their public value proposition. Robust programs and organizations therefore excel at "dynamic conservatism" (Goodsell 2011; Schon 1971).

Many public policies, programs, networks and agencies have very long life-spans. They not only must perform well at any point in time, but also over time, and thus in the face of only partly foreseeable circumstances and changes (Capano and Woo, 2017; Howlett et al., 2018). Climate change governance is a notorious area in which long-term resilience and adaptive capacity are frequently insufficient (Termeer et al., 2017). The Covid-19 pandemic offers a vivid demonstration of the importance of an agile-adaptive response to fast-changing circumstances, with success depending on adaptability and scalability of public health and business support policies to crisis levels (Christensen and Lægreid, 2020). In extreme circumstances, successful governance can be realized overnight—business support programs offered rapidly available income support for businesses after Covid-lockdowns. How well policies adapt over time however is crucial to the endurance of success (Compton and 't Hart, 2019; Luetjens and 't Hart, 2019). Patashnik's (2008) study about what explains the survival of some and the demise of other general interest reforms evolve around the notion of policy (ir)reversibility: how to make sure the core ideas and structural components of a reform package survive the vagaries of the electoral cycle and the variety of sectional interests' lobbies to water it down or wind it back altogether? Altering the composition and identity of the supporting coalition is

pivotal, and Patashnik shows that effective reformers purposefully craft reform ideas, coalitions and policy instruments that help bring this about. Other scholars point to the potential uses of careful institutional layering, bricolage and experimentation for arriving at robust and resilient governance systems (Sabel and Zeitlin 2012; Van der Heijden 2011).

The learning test: Is it continuously working to improve itself?

Is there evidence of effective systems and practices of continuous improvement? Is there a demonstrated record of absorbing changes and surprises whilst maintaining performance and reputation?

Whether put into place through foresight in design or emerging along the way through effective practices of professional, social, and political accountability, the capacity to learn from experience is an essential requirement for the sustainability of governance successes. Goodsell (2011) offers in-depth accounts of the trajectories and governance features underpinning the impressive track records and strong reputations of six U.S. public agencies and catches their commonalities in a 9-cell matrix, one entire row of which is devoted to "temporal aspects"—thus filling the missing link in the Peters and Waterman approach. Goodsell (2011, 14–25) furthermore found three key "sustaining features" supporting an organization's ability to maintain their performance and reputation over time. They are: (a) "beliefs are open to contestation and opposition"—nothing in the organization's make up and practices is ever completely taken for granted and undiscussable; (b) "qualified policy autonomy to permit appropriate change"—front-line professionals and support staff are given a license to do things differently if they think this will lead to improvements or effective response to changing conditions or new demands; (c) "agency renewal and learning are ongoing"—they have the ability to "be innovative but not make a fetish out of it for its own sake" (ibid, 24). In combination, they constitute what we would call learning capacity.

In these successful agencies, Goodsell (2011, 25) observes, "efforts are undertaken to reshape the agency's ethos so that it becomes culturally habituated to dealing with change as an ever-present possibility." We all know how hard this is in a world of rule-driven bureaucracy, hard-fought compromises, and path-dependent policies, but Goodsell's vivid accounts of old yet vibrant and adaptive agencies like the U.S. National Park Service show it can be done. Leadership that provides license to innovate plays an important role. But at the same time, even some "leaderless," transnational and hybrid public-private networks are able to cultivate this quality, as shown by the remarkable institutionalization of a learning culture in the global civil aviation safety regime (SKYbrary, 2017).

Table 1 summarizes the assessment funnel and its constituent tests, and offers key prompting questions for each of the three sets of tests it contains.

Table 1. The assessment funnel: operational framework for case assessment.

Criteria and tests	Operationalization
1. Meaningful public contribution	
1a. The social impact test: Does it add value?	Does the pattern of (intended and unintended) social benefits that results from the initiative outweigh its costs?
	Is this pattern of benefits and costs positively valued across the spectrum of stakeholders?
1b. The delivery test: Does implementation work?	Are the implementation mechanisms based on solid evidence and scientific insight, and adapted to the context in which the activities take place?
2. Prudent use of public power	
2a. The legitimacy test: Is it lawful and just?	Are governance processes and outcomes in accordance with the rule of law and perceived as just and fair by all stakeholders?
2b. The responsiveness test: Is it accountable?	Do the key public actors involved in a governance practice engage in proactive and responsive account-giving to multiple audiences that allow these to be well-informed about and able to evaluate its merits and progress?
3. Sustainable public performance	
3a. The robustness test: Does it perform well over time?	Are considerations of long-term sustainability given due attention in the institutional design and management of the initiative?
3b. The learning test: Is it continuously working to improve itself?	Is there evidence of effective systems and practices of continuous improvement?
	Is there a demonstrated record of absorbing changes, challenges and surprises whilst maintaining performance and reputation?

Using the funnel framework in the classroom

We applied the funnel framework in multiple classroom settings, ranging from bachelor courses to executive workshops and graduate teaching. From these experiences, we drafted a five step approach to using this funnel framework for teaching students about governance success.

Step 1 – students explore the normative content and design philosophy underpinning the framework

Rather than just giving students an evaluation tool to wield, the funnel's transparent normative core of three metacriteria enables teachers to get students to think about the design choices evaluators face in selecting and molding criteria from them into an operationalized assessment framework. In opening sessions of our courses, we typically

situate the framework by rooting the framework in three key traditions of good governance thinking: public value theory (e.g., Moore, 2013), cybernetic theory (Deutsch, 1960; Luhmann, 1995), and procedural justice theory (Tyler, 1990, 2001).

We then invite students to critically interrogate the focus on three among the larger set of normative criteria that inform the eight hallmarks of good governance propagated by the United Nations (participatory, consensus-oriented, accountable, transparent, responsive, effective and efficient, observing the rule of law, equitable and inclusive). This can lead into demonstrations of how choices concerning of evaluation criteria affect the outcomes of evaluations, and therefore what (kinds of) governance success and failures evaluators are primed to see, and not see.

We also use the opportunity to get students to think about the structure of challenges of multicriteria evaluation frameworks. We invite them to consider and challenge the rationale for structuring the framework not as the oft-used traditional multicriteria "web" but as a funnel dictating a step-by-step screening and sorting of cases into categories and degrees of success (i.e., as complete failures, conflicted successes, partial successes, complete successes analogous to McConnell, 2010).

Step 2 – Students apply the framework to the case(s), using high-quality secondary and/or contemporary source materials to inform their assessment

The framework comes to life in application to a body of comparable cases. There are now various open access sources offering hundreds of (more or less) "thick description" case studies of purported public policy, project and collaborative government successes that can be used to find case and background material for the analysis. Lecturers can themselves makes a selection of one or multiple cases from the case collections on offer. Alternatively, students with more methodological grounding, such as graduate students, can first be challenged to come up with a reasoned selection criteria for them to construct a set of comparable cases (selecting on policy sector, jurisdiction/region, political regime type, or historical period, etc.).

Helpful go-to repositories include:

1. The Public Impact Observatory of the Center for Public Impact: https://www.centreforpublicimpact.org/insights
2. The Collaborative Governance Case Database of collaborative governance scholars: https://collaborativegovernancecasedatabase.sites.uu.nl/
3. The John L Alford case library of the Australia New Zealand School of Government: https://www.anzsog.edu.au/resource-library/case-library/the-case-program
4. The cases included in several open access volumes of policy successes hosted by the Successful Public Governance program at Utrecht University's School of Governance: https://www.uu.nl/en/research/successful-public-governance/cases-successful-public-governance (see also Compton and 't Hart, 2019; De La Porte et al., forthcoming; Howlett et al., forthcoming; Luetjens and 't Hart, 2019).

Table 2. Applying the assessment funnel.

Case reviewer	A	B	C	A	B	C	A	B	C
Criteria	Partial success – GI bill (USA)			Conflicted success - petroleum future fund (Norway)			Complete success - HIV/Aids policy (Australia)		
1. Meaningful public contribution creation									
1a. The social impact test: Does it add value?	+++	++	++	++	+++	+++	+++	+++	+++
1b. The delivery test: Does the implementation work?	++	+++	++	++	++	+++	+++	+++	+++
2. Prudent use of public power									
2a. The legitimacy test: Is it lawful and just?	+	+	-	+	++	++	+++	+++	+++
2b. The responsiveness test: Is it accountable?	++	+	0	++	-	+	+++	+++	+++
3. Sustainable public performance									
3a. The robustness test: Does it perform well over time?	++	+++	+++	++	++	+++	++	++	++
3b. The learning test: Is it continuously working to improve itself?	+	+++	++	++	++	++	++	++	++

Note. Scores range on a seven-point scale: +++, ++, +, 0, -, --, ---.

Each of these repositories contains synthetic descriptions and sometimes explicit evaluations of cases, but also includes references to key primary and secondary sources that offer essential routes into deeper insight into the context, actors, processes and outcomes of the policies, organizations, and collaborations in focus. Consulting these extra materials is often necessary for students applying the funnel framework, as the initial case study text may not always offer enough empirical ground to stand on when applying the tests and it maybe enriching to consider multiple perspectives on the case (Mushkat, 2001). Students can capture their assessments on the various criteria in a simple scorecard, as exemplified in Table 2. More importantly, they should accompany these scores with a brief outline of their thought process, to provide the basis for deliberation and comparison in the next steps.

Step 4 – students debate the similarities and differences in their assessment scores, exploiting disagreements to highlight the complexities of evaluating cases.

The funnel framework is a discursive tool, aimed to facilitate systematic reflective discussion about the evaluation of complex cases (Connolly et al., 2015). Precisely because it contains multiple perspectives on "good governance" operationalized into six

specific "tests," the funnel framework helps to identify areas of where reasonable observers may disagree about the balance of evidence for a particular test or indeed about specific cases should be scored.

Consider Table 2, which summarizes the separate evaluations of three reviewers of three policy programs. Beginning at its right-hand column, case reviewers are consistent in their assessment of the HIV/Aids program. All cited the lowest infection rates in the world, the long term saving of billions of healthcare costs by early intervention, and the active engagement with potentially alienated communities as justification for their scores. These three elements form three very different forms of public value, but were all seen to be of equal importance by the three assessors. This consistency of assessments for this case also extended to the prudent use of power (filter 2) and sustainability of the performance (filter 3), again with all reviewers citing similar characteristics of the program in their justifications. In short, the application of the funnel yields uniform support for it being marked as a "complete success."

The application of the Funnel to the GI Bill demonstrates the funnel framework's ability to yield nuanced assessments and to discriminate between complete and partial successes. What has become known colloquially as the "GI Bill" has become widely hailed as one of the standout American public policy achievements of the 20th century (Compton, 2019). Its overall societal impact was remarkable: it lifted levels of education attainment nationwide by 20%, increased home-ownership, and created a "civic generation" that provided the backbone of the US's democratic fabric in the postwar decades.

Yet, when forced to consider the program's distributional impacts and the moral integrity with which it was implemented more closely, students become aware of—and thus to take into account in assessing—critical stains on this success. Though the final bill received broad bipartisan support, part of the legislative compromise was that state governments would be given a key role in administering the scheme. This provided the segregated states with a lever to prevent payments getting to veterans of color, which they duly applied. Likewise, female WW2 veterans were also ill-served by the implementation practices that ensued. There were no checks and balances stopping this from happening: apart from formal Congressional oversight there were no accountability mechanisms. The assessors all note these issues of fairness and legitimacy but reach different conclusions about how to weigh these in their overall scoring on the legitimacy and responsiveness tests. This then allows reflective classroom discussion about the inherent difficulties of assigning values to different criteria in multi-criteria evaluation designs.

The third case concerns Norway's Petroleum Future Fund sparked significant differences between the assessors. When applying filter 1, all duly note its strong public value proposition of using current oil wealth for both short-term and long-term public purposes in a balanced manner that has helped the national economy and the state's finances to effectively dodge the notorious "resource curse" (Ross, 2015). In process terms, they acknowledge that the fund's far-sightedness was the product of the decision to purposely insulate an unusually high proportion of resource income from short-term spending pressures.

The Fund's institutional design puts its administrators at arms-length from political control comparable to other non-majoritarian regulatory and adjudication institutions that

derive both the effectiveness and their social legitimacy from their aloofness. In assessing the legitimacy and responsiveness of the Fund (filter 2) some assessors are more critical than others of the small, technocratic circle of politicians and bureaucrats which founded the fund, arguing that broader support for it was only obtained post-hoc by politicians "selling" the fund well after its creation. It is not until filter 3 comes into play that opinions really start to diverge. One assessor highly rates its long-term vision as prime evidence of its concern with sustainability. Another, however, assigns low scores noting that in its first decade the Fund was not designed to practice sustainable investment policies, thus making money off exploitative activities in other parts of the world. And yet another rates its sustainability as quite low because the fund's core income rests upon the fundamentally unsustainable extraction and distribution of fossil fuels. These diverging assessments on the same criterion can be highlighted in class, to make students aware of the look and feel of a "conflicted success," the political challenges this may bring for policymakers and administrators, and the political opportunities it offers for advocates of change.

Step 5 (Optional) – Students perform process-tracing analysis of selected high-end and low-end cases to reflect on the driving forces and critical enablers of success

The funnel framework is focused exclusively on identifying and assessing cases of success. It does not offer students any levers for the equally complicated challenges that await them when seeking to explain and learn from instances of success in governance. However, its application in the classroom can be used as a launch pad for these questions. Take for example the perennial agency versus structure/context question faced in any explanatory endeavor within the social sciences. Going back to our three exemplary cases, several questions that can be pursued by more in-depth process-tracing reconstructive and analytical work emerge. To what extent can governance successes of the Norwegian Petroleum Fund be attributed to the qualities of the policymakers involved and to what extent did auspicious contextual factors conduce towards policy success? The relatively late discovery of Norway's oil allowed policymakers to be attuned to the hazards of the resource curse that had beset so many other resource-rich countries' economies and political systems. Perhaps it was this awareness that motivated them to grasp relatively early on that the governance of the fund should be firmly embedded in Norway's already mature model of societal corporatism and democratic accountability.

Likewise, the absence of established players and entrenched structures in the HIV/Aids sector presented Australian policy makers with a relatively blank canvas upon which they could design almost from scratch the processes and forge the relationships that would prove so effective in guiding the nation's response to HIV Aids. By contrast, the GI Bill had to be enacted within the US's system of separation of powers and its delicate federal-state relations, necessitating political compromise that left the implementation of the program exposed to racist structures prevalent in some states.

These brief vignettes suggest that moving students from assessing and explaining instances of governance successes requires educators to get students to carefully consider the nature of the hand that policymakers are being dealt by context and circumstance.

They cannot simply—if conveniently—resort to simple agency-centric and post-hoc, propter-hoc explanations along the lines of "good outcomes must be due to good leadership" (see Herek et al., 1987 and Schafer and Crichlow, 2010 for nuanced discussions).

Reflections and implications

The design and application of the funnel framework reflects the conviction that debating the merits of specific instances of public governance should not be limited to an accounting exercise but should reflect the complex nature of governing itself. It should therefore be driven by a broad spectrum of evaluation criteria, and cannot take the form of simple and dichotomous "box-ticking" exercises. It should leave space for nuanced judgment and take shape in deliberative processes. Indeed, the framework's design as a normatively grounded multicriteria assessment tool—forces users to examine and assess any particular case from a broader perspective than classic performance assessment approaches do.

That said, we are acutely aware that both its design reflects choices on our part that are open to disagreement and may have inherent shortcomings. One area of needed development would involve probing whether the three filters are as mutually independent as we present them here. For example, public value scholars such as Mark Moore (1995) argue that the support a public program's or project enjoys among the stakeholders in its "authorizing environment" is an essential precondition for its ability to deliver valued outcomes.

Moreover, the funnel framework could progress from the kind of static one-shot, summative assessments—where in every case a single value is assigned for each test—towards dynamic, multi-shot, and more granular modes of assessment. This would involve teaching students how to "slice" cases into several meaningfully distinct chronological episodes (i.e., waves of policy design, roll-out, adjustment an reform and/or stages in institutional life-cycles) or functional parts (e.g., breaking down a big umbrella policy such as the GI Bill into specific educational, health care, and psychosocial support programs), and to examine whether and how these can be assessed distinctly to arrive at better grounded, more nuanced summative judgments.

Third, students can be encouraged to follow in the footsteps of, for example, Bovens et al. (2001) and examine the impact of taking different temporal vantage points on governance success assessments, for example by comparing assessments for the original GI Bill with that of subsequent amendments, and its later performance during multiple demobilizations following America's various post WW2 wars). Just as different assessments of success over time may emerge from a historical perspective, different geographical and institutional contexts may lead to different assessments of success. The successful governance assessment funnel, its tests, and the applications presented in this article have been developed from the perspective of advanced industrialized western democracies. We assume that the funnel can also be applied in developing countries, and its future development could address its suitability across a variety of settings. One particular question would be if successful policies—including a pass on the legitimacy

test—can develop despite an overall absence of the Rule of Law in authoritarian regimes—in other words can successful initiatives sustainably develop in infertile, anti-democratic soil.

Finally, the funnel framework can also lead us to think deeper about the difference between "good enough" and "great" public governance (cf. Collins, 2001). Perhaps one measure of greatness in public policymaking is about its transformative qualities. Truly successful cases of governance do not just deliver the goods as planned, but *transform* aspirations and values, forge productive relationships and inspire others by demonstrating what is possible. For example, the HIV/Aids program in Australia was groundbreaking in its use of a community-directed, patient-centered approach. It has since been credited by public health scholars for setting a new paradigm for a more egalitarian approach to health care and to doctor-patient relationships. This program was not only able to legitimize its own modus operandi, but provided a beacon for other fields by setting a new standard and a modus operandi that resonated not just within but far beyond its country of origin (Fitzgerald et al., 2019).

On the whole, we think the funnel framework has a meaningful contribution to make to public administration classrooms in courses dealing with policy evaluation, performance measurement, public value management, and institutional learning. Its focus on searching for positives offers a needed counterpoint to the pre-occupation with public sector failures. Its multicriteria design facilitates a more systematic and nuanced perspective on success and failure. And finally, its qualitative-deliberative design helps student to see the important of not just evidence but argument and persuasion in evaluating performance of public sector bodies and programs.

Declaration of conflicting interests

Funding

ORCID iDs

Scott Douglas https://orcid.org/0000-0002-3548-4899
Paul 't Hart https://orcid.org/0000-0003-1454-0932

Note

1. European Parliament resolution of 12 September 2018 on a proposal calling on the Council to determine, pursuant to Article 7(1) of the Treaty on European Union, the existence of a clear risk of a serious breach by Hungary of the values on which the Union is founded (2017/2131(INL)). http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&reference=A8-2018-0250&language=EN#title3

References

Alford J, Douglas S, Geuijen K, et al. (eds) (2017) Special issue: ventures in public value management. *Public Management Review* 19(5): 589–685.

Bach T and Wegrich K (eds) (2018) *The Blind Spots of Public Bureaucracies and the Politics of Non-Coordination*. London, UK: Palgrave.

Behn RD (2001) *Rethinking Democratic Accountability*. Washington, DC: Brookings.

Boardman A, Greenberg D, Vining A, et al. (2017) *Cost-Benefit Analysis: Concepts and Practice*. Cambridge, UK: Cambridge University Press.

Boudon R (2016) *The Unintended Consequences of Social Action*. New York, NY: Springer.

Bovens M and 't Hart P (1996) *Understanding Policy Fiascoes*. New Brunswick, NJ: Transaction.

Bovens M, 't Hart P and Peters BG (eds) (2001) *Success and Failure in Public Governance: A Comparative Analysis*. Cheltenham, UK: Edward Elgar.

Bryson JM, Crosby BC and Bloomberg L (eds) (2015) *Public Value and Public Administration*. Washington, DC: Georgetown University Press.

Capano G and Woo JJ (2017) Resilience and robustness in policy design: a critical appraisal. *Policy Sciences* 50(1): 1–28.

Carpenter DP (2001) *The Forging of Bureaucratic Autonomy: Reputations, Networks and Policy Innovation in Executive Agencies, 1862-1928*. Princeton, NJ: Princeton University Press.

Christensen T and Lægreid P (2020) Balancing governance capacity and legitimacy: how the Norwegian government handled the COVID -19 crisis as a high performer. *Public Administration Review* 80: 774–779.

Coglianese C (2016) *Achieving Regulatory Excellence*. Washington, DC: Brookings.

Collins J (2001) *Good to Great*. New York, NY: Harper Collins.

Comfort LK, Boin A and Demchak C (eds) (2010) *Designing Resilience: Preparing for Extreme Events*. Pittsburgh, PA: Pittsburgh University Press.

Compton M (2019) The 'social warfare state': Americans' making of a civic generation. In: Compton ME and 't Hart P (eds) *Great Policy Successes*. Oxford, UK: Oxford University Press.

Compton M, Douglas S, Fahy L, Luetjens J, 'tP Hart and van Erp J (2022). New development: Walk on the bright side—what might we learn about public governance by studying its achievements? *Public Money & Management*, 42:1, 49-51, DOI: 10.1080/09540962.2021.1975994

Compton M and 't Hart P (eds) (2019) *Great Policy Successes* Oxford, UK: Oxford University Press.

Connolly J, Reid G and Mooney A (2015) Facilitating the evaluation of complexity in the public sector: learning from the NHS in Scotland. *Teaching Public Administration* 33(1): 74–92.

Cristofoli D, Meneguzzo M and Riccucci N (eds) (2017) Special issue - collaborative administration: the management of successful networks. *Public Management Review* 19(3): 275–380.

Cooper T and Wright ND (eds) (1992) *Exemplary Public Administrators*. San Francisco, CA: Jossey-Bass.

De Waal A (2010) Achieving high performance in the public sector: what needs to be done. *Public Performance and Management Review* 34(1): 83–106.

de La Porte C, Eydal G, Kauko J, et al. (eds) *Successful Public Policy: Lessons from the Nordic Countries* (forthcoming). Oxford, UK: Oxford University Press.

Deutsch K (1960) *The Nerves of Government*. New York, NY: Free Press.

Dickinson H and Sullivan H (2014) Towards a general theory of collaborative performance: the importance of efficacy and agency. *Public Administration* 92(1): 161–177.

Douglas S, Schillemans T, 't Hart P, Ansell C, Bøgh Andersen L, Flinders M, Head B, Moynihan D, Nabatchi T, O'Flynn J, Peters BG, Raadschelders J, Sancino A, Sørensen E and and Torfing J (2021) Rising to Ostrom's challenge: an invitation to walk on the bright side of public governance and public service. Policy Design and Practice 4(4): 441-451. doi: 10.1080/25741292.2021.1972517.

Fetterman DM and Wandersman A (2005) *Empowerment Evaluation Principles in Practice*. New York, NY: Guilford Publications.

Fetterman D (2006) *Foundations of Empowerment Evaluation*. New York, NY: Sage.

Fischer F (1995) *Evaluating Public Policy*. Chicago, IL: Nelson-Hall.

Fitzgerald L, Mutch A and Herron L (2019) Responding to HIV/AIDS: mobilisation through partnerships in a public health crisis. In: Luetjens J, Mintrom M and 't Hart P (eds) *Public Policy Successes in Australia and New Zealand*. Canberra, Australia: ANU Press.

Goodsell CT (2011) *Mission Mystique: Belief Systems in Public Agencies*. Washington, DC: CQ Press.

Gray P and 't Hart P (1998) (eds) *Public Policy Disasters in Europe: A Comparative Analysis*. London, UK: Routledge.

Hall P (1981) *Great Planning Disasters*. Berkeley, CA: University of California Press.

Hartley J, Sørensen E and Torfing J (2013) Collaborative innovation: a viable alternative to market competition and organizational entrepreneurship. *Public Administration Review* 73(6): 821–830.

Herek GM, J anis IL and Huth P (1987) Decision making during international crises: is quality of process related to outcome? *Journal of Conflict Resolution* 31(2): 203–226.

Hinterleitner M (2018) Policy failures, blame games, and changes to policy practice. *Journal of Public Policy* 16(1): 81–106.

Hood C (1991) A public management for all seasons? *Public Administration* 69(1): 3–19.

Howlett M, Mukherjee I and Woo JJ (2015) From tools to toolkits in policy design studies: the new design orientation towards policy formulation research. *Policy & Politics* 43(2): 291–311.

Hood C (2010) *The Blame Game*. Princeton: Princeton Univerity Press.

Howlett M, Capano G and Ramesh M (2018) Designing for robustness: surprise, agility and improvisation in policy design. *Policy and Society* 37(4): 405–421.

Howlett M, Lindquist EM and Skogstad G. In: Tellier G and 't Hart P (eds) *Successful Public Policy: Lessons from Canada* (Forthcoming). Oxford, UK: Oxford University Press.

Jennings W, Lodge M and Ryan M (2018) Comparing blunders in government. *European Journal of Political Research* 57(1): 238–258.

King A and Crewe I (2013) *The Blunders of Our Governments*. London, UK: Oneworld.

Light P (2014) *A Cascade of Failures: Why Government Fails, and How to Stop it*. Washington, DC: Brookings Center for Effective Public Management.

Lindquist E and Marcy R (2016) The competing values framework: implications for strategic leadership, change and learning in public organizations. *International Journal of Public Leadership* 12(2): 167–186.

Luetjens J., Mintrom M. and Hart 'tP. (eds) (2019) *Successful Public Policy: Lessons From Australia and New Zealand*. Canberra: ANU Press.

Luetjens J. and 't Hart P. (2019) *Governing by Looking Back: Learning from Successes and Failures*. Australian Public Service Review Panel.

Luhman N (1995) *Social Systems*. Stanford, CA: Stanford University Press.

Maguire ER (2018) New frontiers in research on procedural justice and legitimacy in policing. *Police Practice and Research* 19(2): 107–110.

McConnell A.(2010) Success, failure, or somewhere in-between? A framework for evaluating crisis management. *Policy and Society* 30(2): 63–76.

McDonald BD III (2021) Teaching in uncertain times: the future of public administration education. *Teaching Public Administration* 39(1): 3–8.

Miller WR, Sorensen JL, Selzer JA, et al. (2006) Disseminating evidence-based practices in substance abuse treatment: a review with suggestions. *Journal of Substance Abuse Treatment* 31(1): 25–39.

Moore M (1995) *Creating Public Value: Strategic Management in Government*. Cambridge, MA: Harvard University Press.

Moore M (2013) *Recognizing Public Value*. Cambridge, MA: Harvard University Press.

Mulgan R (2014) *Making Open Government Work*. London, UK: Bloomsbury Publishing.

Mushkat M (2001) Teaching case analysis to public managers in academic settings. *Teaching Public Administration* 21(1): 37–45.

Nielsen SB, Turksema R and van der Knaap P (eds) (2015) *Success in Evaluation: Focusing on the Positives*. New Brunswick, Canada: Transaction.

Nutley SM, Walter I and Davies. HTO (2007) *Using Evidence: How Research Can Inform Public Services*. Bristol, UK: Policy Press.

Opperman K and Spencer A (eds) (2016) Fiascos in public policy and foreign policy. *Journal of European Public Policy*, 23(5): 643–787.

Page SB, Stone MM, Bryson JM, et al. (2015) Public value creation by cross-sector collaborations: a framework and challenges of assessment. *Public Administration* 93(3): 715–732.

Patashnik E (2008) *Reforms at Risk*. Princeton, NJ: Princeton University Press.

Prebble M (2018) Is "We" Singular? The Nature of Public Value. *The American Review of Public Administration* 48(2): 103–118. doi: 10.1177/0275074016671427.

Rochlin G (1996) New directions in reliable organization research. *Journal of Contingencies and Crisis Management* 4(2): 55–111.

Roe E and Schulman P (2008) *High-Reliability Management*. Stanford, CA: Stanford University Press.

Rosenbloom DH (1983) Public administrative theory and the separation of powers. *Public Administration Review* 43(3): 219–227.

Ross ML (2015) What have we learned about the resource curse? *Annual Review of Political Science* 18(2): 239–259.

Rothstein B (2012) Political legitimacy for public. In: *The Sage Handbook of Public Administration*. London, UK: Sage, 407–419.

Sabel CF and Zeitlin J (2012) Experimentalist governance. In: Levi-Faure D (ed) *Oxford Handbook of Governance*. Oxford, UK: Oxford University Press. DOI: 10.1093/oxfordhb/9780199560530.013.0012.

Schafer M and Crichlow S (2010) *Groupthink vs High Quality Decision Making in International Relations*. New York, NY: Columbia University Press.

Schon D (1971) *Beyond the Stable State*. New York, NY: Random House.

SKYbrary (2017) *Aviation Safety Performance Reports and Statistics*. Available at: www.skybrary.aero/index.php/Aviation_Safety_Performance_Reports_and_Statistics (accessed 8 February 2018).

Sørensen E (2017) Political innovations: innovations in political institutions, processes and outputs. *Public Management Review* 19(1): 1–19.

Suchman MC (1995) Managing legitimacy: strategic and institutional approaches. *Academy of Management Review* 20(3): 571–610.

Tamanaha B (2007) A Concise Guide to the Rule of Law. *Legal Studies Research Paper Series #07-0082*. Jamaica, NY: St John's University School of Law. Available at: http://content.csbs.utah.edu/~dlevin/conlaw/tamanaha-rule-of-law.pdf (accessed 28 September 2021)

Termeer C, Van Buuren A, Dewulf A, et al. (2017) Governance arrangements for adaptation to climate change. In: *Oxford Research Encyclopedia of Climate Science*. Oxford, UK: Oxford University Press. DOI: 10.1093/acrefore/9780190228620.013.600.

Tyler TR (1990) *Why People Obey the Law*. New Have, CT: Yale University Press.

Tyler TR (2001) A psychological perspective on the legitimacy of institutions and authorities. In: Jost J. and Major B. (eds) *The Psychology of Legitimacy*. Cambridge, UK: Cambridge University Press.

Van Assche K, Beunen R and Duineveld M (2011) Performing success and failure in governance: Dutch planning experiences. *Public Administration* 90(3): 567–581.

Van der Heijden J (2011) Institutional layering: a review of the use of the concept. *Politics* 31(1): 9–18.

Waldron J (2016) The rule of law. In: Zalta E (ed) *The Stanford Encyclopedia of Philosophy*. Available at: https://plato.stanford.edu/archives/fall2016/entries/rule-of-law/(accessed 28 September 2021)

Walker B and Salt D (2006) *Resilience Thinking: Sustaining Ecosystems and People in a Changing World*. Washington, DC: Island Press.

Weick K and Sutcliffe K (2011) *Managing the Unexpected: Resilient Performance in an Age of Uncertainty*. New York, NY: Wiley.

World Health Organization (2017) Assessment Report of the Global Vaccine Action Plan. Available at: https://www.who.int/immunization/web_2017_sage_gvap_assessment_report_en.pdf (accessed 28 September 2021)