Searching for successful governance: a heuristic assessment tool

Scott Douglas, Judith van Erp, Paul ‘t Hart

EGPA 2018

Abstract

Politicians, journalists, citizens, and scholars excel in pinpointing and dissecting the failures and shortcomings of government. Considerably less attention is paid to instances of governments performing well. To address this imbalance and further our understanding of the nature of successful public governance, we need to grasp what ‘success’ means and how it is attained. This article proposes an assessment framework designed to systematically identify and analyze cases of successful governance. Set up as a funnel consisting of three assessment filters, each containing specific tests, the framework operationalizes governance success in terms of (1) meaningful public outcomes, (2) the prudent use of public power, and (3) the ability to sustain public performance over time. Each of the three filters will be elaborated in light of existing literature. We illustrate how the framework can operate through applying it to three cases: the American GI Bill, the Norwegian Oil Fund, and the Australian HIV/Aids intervention program.

Keywords: Positive Public Administration, performance, public value, legitimacy, accountability, sustainability, endurance

Word count (excluding appendices and tables, including references): 8.280

Debating Governance: Changing the Frame

Public governance scholarship has in recent decades built up a rich language persuading us just how difficult it is to govern well, particularly in late-modern conditions. We are now routinely told that wicked problems, 'VUCA' (volatility, uncertainty, complexity and ambiguity), risk, disturbances and crises are the new normal in both the operational and the authorizing environments of contemporary governments (Van der Wal 2017; Laegreid and Rykkja 2019). Moreover, the expansion and institutionalization of a wide array of transparency and accountability arrangements (Bovens et al. 2014; Mulgan 2014) have paradoxically resulted in much energy being directed to pinpointing and dissecting cases in which governments fail our expectations, and to governments organizing their defenses against such charges. The contemporary mood of 'declinism' in the West has led to a resurgence in the use of negative metaphors of illness, breakdown, crisis, collapse, decay and death to characterize the state of our political systems and our public institutions (Diamond 2005; Fukuyama 2014; Levitsky and Ziblatt 2018). By now, there is a vast body of speeches, news items, investigation reports, and scholarly studies on government 'disasters' (Hall 1981; Gray and 't Hart 1998), blunders (King and Crewe 2013; Jennings, Lodge, and Ryan 2018), policy failures (Bovens and 't Hart 1996; Light 2014; Opperman and Spencer 2016), blind spots (Bach and Wegrich 2018), and blame games (Hood 2010; Hinterleitner 2018). Studying these 'dark sides' is important. It can teach us what to avoid, prevent and contain when crafting public institutions, positioning regulatory agencies, designing and implementing public policies, and managing large public projects and innovation processes.

Unfortunately, our current public discourses and academic practices concerning governance are not equally attuned to spotting and naming successes as they are to finding faults and blaming public officials and agencies for them. Well-performing public institutions and policies are what many people in advanced economies and established democracies take for granted: that power changes hands peacefully, that public health, safety and security are effectively regulated and policed, that the rule of law is unquestioned, that government institutions are incorruptible, that there is trust and confidence in the economic order, and that public services are generally delivered effectively, efficiently and fairly. If the study of governance failure and breakdown teaches us what to avoid, studying governance success can teach us what to embrace, aim for and emulate.

In this article we seek to address this imbalance by presenting a framework for assessing and analyzing cases of successful governance. 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 balanced examination of what success in governance looks like and how it comes about. It is designed to be directly applicable as a parsimonious and versatile assessment tool that can be applied to different types of governance arrangements and cases (policies, programs, projects, networks) and incorporated into different types of research designs.

In presenting the framework, we seek to both build upon and transcend existing foundational work on positive policy evaluation (Fetterman and Wandersman 2005; Fetterman 2006; Nielsen, Turksema, and van der Knaap 2015); policy success (McConnell 2010; Compton and ‘t Hart 2019; Luetjens et al. 2019); regulatory excellence (Coglianese 2016); public value creation (Moore 1995 2013; Bryson et al. 2015; Alford et al. 2017); successful collaborative governance and network management (Dickinson and Sullivan 2014; Page et al. 2015; Cristofoli , Meneguzzo, and Riccucci 2017); political and public innovation (Hartley, Sørensen, and Torfing 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 (Walker and Salt 2006; Comfort, Boin, and Demchak 2010); and high-reliability systems performing public tasks in high-risk operating environments (Rochlin 1996; Roe and Schulman 2008; Weick and Sutcliffe 2011).

Much of this research generates valuable insight into ‘what works’ across different functions, tasks, sectors and levels of governance. Yet these researchers vary widely in how they choose to conceptualize and assess desirable attributes such as high performance, high effectiveness, high public value, high reliability - in short: success. This paper aims to provide a unifying language, embodied in an assessment tool, for doing so.

Assessing Governance Success: The Funnel Framework

In constructing the heuristic framework we make certain assumptions about the nature of successful governance that are important to articulate up front. First, we focus on 'cases of governance' which we define as discrete bundles of activity aimed at addressing public issues, undertaken by government and/or in collaborations between government and other social actors. Governance can take the shape of projects, programs, collaborations, and partnerships in a variety of contexts. Further, we assume governance success is *multidimensional* (it can, and needs to be, assessed from multiple vantage points and value sets), *constructed* (in the political world of government and governance, terms like 'success', 'failure', 'performance' and 'value' send signals, are inherently selective, are used as weapons, shields and trophies – in short, they are performative, Van Assche, Beunen, and Duineveld 2011), *temporally and culturally contingent* (assessments vary change over time as both the criteria or weight given to them change and as new information about processes and outcomes becomes available), and *situationally contingent* (success has different shapes and drivers in different contexts).

Given this complex nature of 'success', how can we identify cases of governance that display different levels of it, for comparative-analytical and/or practical purposes? While we acknowledge the boldness of even asking the question in such general terms, we surmise that any 'case of governance' can be considered completely successful when it fully satisfies three meta-criteria:

(1) meaningful public contribution of valuable and valued societal results

(2) prudent and legitimate use of public power

(3) sustained performance over time

These three criteria 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), though they derive from underlying normative traditions (Bovens et al. 2008; Meijer, 't Hart, and Worthy 2018). As we will elaborate below, they are constructed as a multi-stage funnel with each stage comprising as a set of assessment filters. We surmise that to be considered a full success, a particular case of governance should be highly rated at all three stages of the assessment funnel process.

FIGURE 1 ABOUT HERE

Figure 1 presents a conceptual map of the framework. The three-fold sequence of 'tests' has been designed to be broadly applicable across a wide range of governance forms and activities. We are well aware that the tests need to be further operationalized depending on the specific features of the case and its political, institutional and operational context. Further below we shall demonstrate what this might look like, but for now it is important to put the spotlight on the construct as a case-identification tool in the broader toolkit of positive governance studies. To this end, we now introduce each of the meta-criteria, the criteria ('tests') subsumed under them, and operationalize these into assessment questions to be used in assessing cases (or sets and populations of cases) of public governance. Table 1 further below provides a synthesis of the framework that can be used for case assessment purposes.

Filter 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?

The first step is to establish whether a particular case of governance adds value to society. This assessment requires a collective costs-benefit analysis of the case of governance evaluated, as is often done for infrastructure, healthcare, and environmental management initiatives (Boardman et al. 2017). Public governance should ultimately expand the aggregate social welfare function, i.e. more benefits and fewer costs for society as a whole. This contribution can be achieved through multiple routes, be it enabling greater effectiveness through cross-societal collaborations or greater efficiencies through enabling technological or social innovation. Either way, this assessment of costs and benefits needs to include the immaterial costs of government action – such as forcing individuals to buy into collective health insurance (Moore 2013), and also cover the unintended consequences of public action – such as the environmental impact of infrastructure projects (Boudon 2016).

However, it may be difficult if not impossible to implement changes which benefit everyone (Arrow 2012). This test should therefore also include an assessment of the distribution of costs and benefits across different actors and stakeholders an important question. Synoptic cost-benefit analysis may struggle to capture the different experiences involved and therefore needs to be complemented with a more contextualized, stakeholder-oriented approach. De Jong (2015) offers a useful framework for identifying the values sought by different stakeholders, distinguishing between material (sustenance, physical health) and non-material (liberty, happiness) values. Such as an exercise reveals to what extent the outcome is considered worthwhile by the stakeholders affected while also gauging 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 pain of the few can outweigh the gain of the many. The assessment must include multi-actor perspectives and differentiate between governance cases where there are universal public or mutual gains and those where costs and benefits are unequally distributed across stakeholder groups.

1b. 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 is designed and service delivery set 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, Walter and Davies 2007). For example, a program fighting substance abuse should be informed by the latest insights in what interventions work best to reduce addiction and relapse (Miller et al. 2006). Also, research on coproduction demonstrates that policy makers can make case by case choices to work through private companies or community groups if that better suits the nature of the task and the make-up of the sector (Alford and Hughes 2008). 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 (WHO 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. And 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 behaviours of different target populations (Howlett, Mukherjee, and Woo 2015). In other words, this test assesses whether the governance case uses implementation strategies and tools which heed these insights and provide the best possible shot at bringing about successful outcomes, taking into account in the specific features of the context at hand.

Filter 2 - Prudent use of public power

2a. 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 wilfulness 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. This is a 'thin' conception of the Rule of Law and excludes substantial requirements, such as human rights, justice, sustainability and social equity (Waldron 2016). And yet Rule of Law demands are 'thicker' than the criterion of legality, as is illustrated by the resolution by the European Parliament stating that

Hungarian laws to curtail judicial independence and freedom of expression are in conflict with the Rule of Law[1].

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.

The norm that the law should apply equally to all does not mean that all are equal. Is a social housing policy successful when it gives new asylum seekers priority over regular tenants? Are individual tax rulings by the tax office for specific corporations justified? These are normative questions demonstrating that substantive values such as justice and fairness and human rights are necessary in addition to the Rule of Law perspective. 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 (Suchman 1995; Page et al. 2015). Legitimacy does not just emerge but instead is actively crafted and developed in processes of legitimation (Suchman 2011; Van Assche , Beunen, and Duineveld 2011). Multiple 'publics' with heterogeneous and potentially conflicting beliefs, values and interests exist to be convinced of the legitimacy of

[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

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 (Tyler 2001; Maguire 2018).

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?

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 (Lecy et al. 2013). 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 dialogue, 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

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?

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? Timing – when to assess – and time horizons are critical in any evaluation. 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. How well they adapt over time is crucial to the endurance of success (Compton and 't Hart 2019; Luetjens et al. 2019). Hence two final subtests are added to the assessment framework.

The Robustness Test takes its cue from Collins and Porras's (1999) classic account of eighteen 'visionary' companies that had consistently outperformed the control group of eighteen comparable and still quite successful companies. A key theme in their analysis was the former group's pre-occupation with their long-term viability. At its core lies the value that is being placed on crafting narratives, methods and commitment that underpin the sustainability of an organization's performance. The preoccupation is to have the will, resources and support keep going when others are thrown off course by changing operating and political environments. Institutionalised organisations have a clear sense of their long-term direction but are adaptable in how they get there in the here and now. They excel at dynamic conservatism (Schon 1971; Goodsell 2011).

In a similar vein, Patashnik's (2008) study about what explains the survival of some and the demise of other general interest reforms evolves around the notion of irreversibility: 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 (Van der Heijden 2011; Sabel and Zeitlin 2012).

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 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 the sustainability 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. In combination, they constitute what we would call learning capacity. 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 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 cultural 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 institutionalisation of a learning culture in the global civil aviation safety regime (www.skybrary.aero 2017).

TABLE 1 ABOUT HERE

Applying the framework: An illustration

We want to demonstrate how the assessment funnel heuristic can be applied to assess, select and compare governance cases by applying it to three well-studied cases. We have selected three chapter length accounts of cases of purportedly successful public governance gleaned from two recent collections of case histories of policy successes (Compton and 't Hart 2019; Luetjens et al. 2019): (1) the United States WWII-era 'GI bill' seeking to improve the country's handling of large numbers of returning servicemen from the current and future wars from the appalling failure to do so following the end of WWI (1945-present); (2) the setting up and management of Norway's Petroleum Future Fund, which stores proceeds of current oil revenues for future generations and the largest sovereign fund in the world (USD 1.000 billion in April 2018) (1989-present); and (3) Australia's response to the

HIV/Aids epidemic which resulted in one of the world's lowest infection rates while advancing minority rights by actively working with the gay community (1980-present).

This small n, purposeful sample of cases selected on the 'dependent variable' – a mortal sin according to most methodologists, but perfectly useful for our present, purely illustrative purpose - should provide a good test of the funnel's ability to provide systematic, differentiated, intersubjectively validated case assessments. If it can do so for these three pre-selected 'success cases', it follows that it will certainly have that ability for any set of randomly selected cases of governance, provided the materials used as sources contain sufficient information to make reasonable assessments for each of the seven tests.

Each of the three authors has separately assessed these three cases using the framework above. Each of the seven criteria was rated on a seven-point scale or marked as 'insufficient data to score'. The scores varied from an extremely negative result (---), strongly negative (--), negative (-), neutral (0), positive (+), strongly positive (++), and extremely positive (+++). We opted for a seven-point scale as we wanted to provide enough bandwidth to distinguish between the different degrees of success. Each author provided short motivation notes on all of their scores for each of the cases, which formed the basis for our subsequent discussion. The Appendix contains the scores and arguments for all of the scores.

Based on our experience in applying the funnel, we rephrased some of the tests to avoid overlap and clarify the meaning of some tests. Most importantly, the distinction between the social impact test, delivery test, and public support test were further specified to refer to the outcomes, implementation, and social support for cases of governance. After our first coding exercise, we slightly adapted the phrasing of some of the tests and subsequently re-coded the cases based on the final version of the funnel. The funnel that was presented above is the version that includes these refinements.

Why bother with this framework? First of all, its design reflects a particular conviction about evaluation and performance measurement that one may share or disagree with, but that we argue has enduring relevance within the field of public administration and management. The core of this conviction is that assessing whether 'taxpayer money is well spent' – in other words 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.

And indeed the design of the funnel framework – a normatively grounded multicriteria assessment tool - forces assessors to obtain a more comprehensive picture of a case than classical evaluation approaches, which are more often than not limited to goal achievement (effectiveness) and efficiency. For example, in case of the Norwegian Petroleum Fund, the assessors had to go back to the original case authors to obtain more information on the Fund's investment policies - opening up the discussion about the 1990s investments in land mines and child labor tainted firms - and the extent to which they were subject to democratic oversight.

Precisely because it contains multiple perspectives on 'good governance' operationalized into seven 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. As evident from table 2, the case reviewers were remarkably consistent in their assessment of the public value outcomes of the HIV/Aids program and the processes through which these were achieved. 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 (see Appendix). 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.

TABLE 2 ABOUT HERE

The application of the Funnel to the GI Bill demonstrates its ability to yield nuanced assessments even for arguably one of the most celebrated American public policy achievements of the 20th century (Compton 2019). Whilst broadly agreeing on its considerable merits, assessors all highlight similar concerns and at the same time diverge in how they weigh these downsides against the upsides of the policy in their final assessment. The public contribution is highly rated across the board, as is the sustainability of this 70-year old policy, although assessors also note that later incarnations of the policy deliver less value to veterans. Yet concerns are raised about the legitimacy of the policy. Though final bill receiving 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.

The case of the Norwegian Petroleum Fund sparked most debate between the assessors. All duly note its 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)(filter 1). But assessors differ markedly about the legitimacy and responsiveness of the Fund (filter 2). Assessor B was critical 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. Assessor A by contrast argues that sovereign wealth funds are purposely created to insulate resource income from political and democratic short-termism. Its very distance from direct political control is actually core to its purpose and legitimacy, similar to the position of other semi-independent bodies. Assessor C shares some of these concerns but reasons that the fund was created in line with the formal democratic systems of checks and balances.

In addition to triggering systematic and transparent debate about what good governance and about governance success looks like within and across concrete cases, the funnel framework allows analysts to localize relative strengths and weaknesses even within cases that at a general summative level rate as 'successful'. Figure 2 provides an illustration of this. The spider webs display the average scores of three cases on each of the seven criteria and together provide a helpful at-a-glance visualization of both the big picture and the specifics of the assessment exercise. It shows, for example, that the GI Bill scores highly on social contribution (1a, 1b) and sustainability (3a, 3b), but much less strongly on prudent use of power (2a, 2b), the Fund's key strengths lie in its social impact (1a) and endurance (3a), whereas the HIV/Aids programs represents the most consistently successful case of the three.

FIGURE 2 ABOUT HERE

Thirdly, the assessment framework enables analysts to usefully compare policies of very different in scope and purpose. The HIV/Aids program, for all its profound impact on the health hazards run by gays and other at-risk groups in the Australian community, arguably had a smaller overall societal impact than the GI Bill which lifted levels of education attainment nationwide by 20%, increased home-ownership, and created a 'civic generation' that provided the backbone of the US's national fabric in the postwar decades. In contrast, the HIV/Aids program respected and enhanced the civic rights of ostracized groups, whereas the implementation of the GI Bill was blighted by biased and discriminatory practices. While the GI Bill may have had a bigger overall impact, the success of the HIV/Aids program is more well-rounded.

In sum, we believe the funnel framework can usefully perform at least two functions for students of governance who are interested in studying the nature and preconditions of successful policies, programs and project in the public sector. First, it can be applied as a heuristic case selection tool for researchers wanting to construct samples for systematic comparisons of successful and less or altogether unsuccessful cases of governance. Second, the funnel has diagnostic potential in that it offers a relatively parsimonious and yet versatile tool to identify different types and levels of success across a number of key dimensions, setting the stage for more granular explanatory analyses than the typical study of government successes (Light 2014; Compton and 't Hart 2019) or failures (King and Crewe 2013; Jennings, Lodge, and Ryan 2018) tends to offer.

Reflections and implications

We hope to have persuaded readers that the funnel framework has a meaningful contribution to make to a self-conscious analytical toolset geared towards identifying, examining, comparing and contrasting, and learning from instances of successful public governance. At the same time, we are acutely aware that both its design and the illustrative application as reported above reflect choices on our part that are open to disagreement and may have inherent shortcomings. We certainly welcome such debate, and hope it will lead to further development of a purposeful eye on and a language for success that our field is currently largely lacking (Douglas et al. 2019).

How to move on from here? One area of needed development would involve probing whether the three filters and seven tests are as mutually independent as we present them here. For example, scholars such as Mark Moore (1995; 2013) - who is on his own and highly influential journey to develop a meaningful methodology of assessing the public value generated by acts of public managers, government agencies and the networks in which they participate - argue that a program's or project's legitimacy among stakeholders is an essential precondition for its ability to deliver valued outcomes. Fuzzy-set comparative case studies could be undertaken to identify whether the presence of high scores on some tests are necessary or sufficient preconditions for high scores on other tests.

Second, future research should seek to move from identifying and describing towards systematically explaining different types and levels of success in governance. As it does so, applications of the funnel framework should progress from the kind of static one-shot, summative assessments – where in every case a single value is assigned for each test – as performed in this article, towards dynamic, multi-shot and more granular modes of assessment. This would involve 'slicing' cases into several meaningfully distinct functional parts (e.g. a series of educational, health care and psychosocial support programs within the larger policy framework of the GI Bill) or temporally distinct episodes (e.g. comparing assessments for the original GI bill with that of subsequent amendments, and its performance during multiple demobilizations following America's various post WW2 wars), and putting each of these individual slices through the funnel assessment process and yielding a series of spider webs and narratives. Taking this step would bring dual benefits. Methodologically it helps us by increasing the all-important 'n' in focused-comparison and process-tracing based research designs (King King, Keohane, and Verba 1994). Substantively it helps identify variations in within-case governance performances that then raise targeted how/why-questions, providing tougher tests for putative explanatory frameworks.

Thirdly, even our current heuristic application of the framework to the three cases discussed yields interesting questions for such a more explanatory research agenda. Take for example the perennial agency versus structure/context question. To what extent can governance successes of the Norwegian Petroleum Fund and the Australian HIV/Aids be attributed to the qualities of the policymakers involved and to what extent did auspicious contextual factors conduce towards policy success? For example, 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 fully matured, 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 both assessments and explanations of governance successes need 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 explanations of the 'good outcomes are due to good leadership', 'bad outcomes are due to bad leadership' variety.

Finally, moving beyond the confines of explanatory research, the funnel framework can also lead us to think deeper about the difference between 'just good' and 'really great' public governance (cf. Collins 2001). Perhaps one measure of greatness in public policymaking is about its transformative qualities. Truly successful cases of governance don't 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). Similarly, the GI Bill not only set a highly influential precedent for the targeted use of direct individual payments by government, but through its investment in college education is also credited by experts for creating the 'civic generation' that went on to become the social backbone for the strength of American democracy (Compton 2019). Great policy successes in other words can serve as beacons and game-changing public-value multipliers, altering for the better the societal and institutional contexts in which they originated.

References

Alford, John, Scott Douglas, Karin Geuijen and Paul ‘t Hart. Eds. 2017. “Special Issue: Ventures in Public Value Management.” *Public Management Review* 19 (5): 589-685.

Alford, J., and O. Hughes. 2008. “Public Value Pragmatism as the Next Phase of Public Management.” *The American Review of Public Administration* 38 (2): 130-148.

Arrow, Kenneth J. 2012. *Social Choice and Individual Values*. New Haven: Yale University Press.

Bach, Tobias, and Kai Wegrich. Eds. 2018. *The Blind Spots of Public Bureaucracies and the Politics of Non-Coordinatio*n. London: Palgrave.

Behn, Robert D. 2001. *Rethinking Democratic Accountability*. Washington: Brookings.

Boin, Arjen, Allan McConnell, and Paul ‘t Hart. Eds. 2008. *Governing After Crisis*. Cambridge: Cambridge University Press.

Boardman, Anthony, David Greenberg, Aidan Vining, and David Weimer 2017. *Cost-Benefit Analysis: Concepts and Practice*. Cambridge University Press.

Boudon, Raymond 2016. *The Unintended Consequences of Social Action*. New York: Springer.

Bovens, Mark, and Paul ‘t Hart. 1996. *Understanding Policy Fiascoes*. New Brunswick: Transaction.

Bovens, Mark, Thomas Schillemans and Paul ‘t Hart. 2008. “Does Public Accountability Work? An Assessment Tool.” *Public Administration* 86 (1): 225-242

Bovens, Mark, Robert Goodin, and Thomas Schillemans. Eds. 2014. *Oxford Handbook of Public Accountability*. Oxford: Oxford University Press.

Bryson, John, Barbara Crosby, and Laura Bloomberg. Eds. 2015. *Public Value and Public Administration*. Washington: Georgetown University Press.

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

Coglianese, Cary. 2016. *Achieving Regulatory Excellenc*e. Washington: Brookings.

Collins, Jim 2001. *Good to Great*. New York: Harper Collins.

Collins, Jim, and James Porras. 1999. *Built to Last*. New York: Harper Business.

Comfort, Louise K., Arjen Boin and Chris Demchak. Eds. 2010. *Designing Resilience: Preparing for Extreme Events*. Pittsburgh: Pittsburgh University Press

Compton, Mallory. 2019. "The 'Social Warfare State': Americans' Making of a Civic Generation." In: Mallory Compton and Paul 't Hart. Eds. 2019. *Great Policy Successes*. Oxford: Oxford University Press.

Compton, Mallory and Paul 't Hart. Eds. 2019. *Great Policy Successes*. Oxford: Oxford University Press.

Cristofoli, Daniela, Marco Meneguzzo, and Norma Riccucci. Eds. 2017. "Special Issue - Collaborative Administration: The Management of Successful Networks." *Public Management Review* 19 (3): 275-380.

Cooper, Terry, and N. Dale Wright. Eds. 1992. *Exemplary Public Administrators*. San Francisco: Jossey-Bass.

De Waal, André. 2010. "Achieving High Performance in the Public Sector: What Needs to be Done." *Public Performance and Management Review* 34 (1): 83-106.

Diamond, Jarred. 2005. *Collapse: How Societies Choose to Fail or Succeed.* New York: Viking.

Dickinson, Helen, and Helen Sullivan. 2014. "Towards a General Theory of Collaborative Performance." *Public Administration* 92 (1): 161-177.

Douglas, Scott, Paul 't Hart, Chris Ansell, Lotte Bogh Andersen, Matthew Flinders, Brian Head, Tina Nabatchi, Janine O'Flynn, B. Guy Peters, Jos Raadschelders, Eva Sørensen, Alessandro Sancino, Thomas Schillemans, and Jacob Torfing. 2019. *Towards Positive Public Administration: A Manifesto.* (Paper under review)

Fetterman, David, and Abraham Wandersman. 2005. *Empowerment Evaluation Principles in Practice*. New York: Guilford Publications.

Fetterman, David. 2006. *Foundations of Empowerment Evaluation*. New York: Sage.

Fischer, Frank. 1995. *Evaluating Public Policy*. Chicago: Nelson-Hall.

Fitzgerald, Lisa, Allyson Mutch, Lisa Herron. 2019. "Responding to HIV/AIDS: Mobilisation Through Partnerships in a Public Health Crisis." In: Jo Luetjens, Michael Mintrom and Paul 't Hart Eds. *Public Policy Successes in Australia and New Zealand*. Canberra: ANU Press.

Fukuyama, Francis. 2014. *Political Order and Political Decay*. Stanford: Stanford University Press.

Goodsell, Charles. 2011. *Mission Mystique: Belief Systems in Public Agencies*. Washington: CQ Press.

Gray, Pat, and Paul ‘t Hart Eds. 1998. *Public Policy Disasters in Europe: A Comparative Analysis*. London: Routledge.

Hall, Peter. 1981. *Great Planning Disasters*. London: Penguin.

Hartley, Jean, Eva Sørensen, Jacob Torfing. 2013. “Collaborative Innovation: A Viable Alternative to Market Competition and Organizational Entrepreneurship.” *Public Administration Review* 73 (6): 821-30.

Hinterleitner, Markus. 2018. “Policy Failures, Blame Games, and Changes to Policy Practice.” *Journal of Public Policy* 16 (1): 81-106.

Hood, Christopher. 1991. “A Public Management for All Seasons?” *Public Administration* 69 (1): 3-19.

Howlett, Michael, Ishani Mukherjee, Jie-Jie Woo. 2015. “From Tools to Toolkits in Policy Design Studies: The New Design Orientation Towards Policy Formulation Research.” *Policy and Politics* 43 (2): 291-311

Jennings, Will, Martin Lodge, and Matt Ryan. 2018. “Comparing Blunders in Government.” *European Journal of Political Research* 57 (1): 238-58.

King, Anthony, and Ivor Crewe. 2013. *The Blunders of Our Governments*. London: Oneworld.

King, Gary, Robert Keohane, and Sydney Verba. 1994. *Designing Social Inquiry*. Princeton: Princeton University Press.

Laegreid, Per, Lisa Rykkja Eds. 2019. *Societal Security and Crisis Management: Governance Capacity and Legitimacy*. London: Palgrave.

Light, Paul. 2014. *A Cascade of Failures. Why Government Fails, and How to Stop It*. Washington: Brookings Center for Effective Public Management.

Lindquist, Evert, and Richard Marcy 2016. “The Competing Values Framework: Implications for Strategic Leadership, Change and Learning in Public Organizations.” *International Journal of Public Leadership* 12 (2): 167-186.

Levitsky, Steven, and Daniel Ziblatt. 2018. *How Democracies Die*. New York: Penguin Viking.

Maguire, Edward. 2018. “New Frontiers in Research on Procedural Justice and Legitimacy in Policing.” *Police Practice and Research* 19 (2): 107-110.

Meijer, Albert, Paul ‘t Hart, Benjamin Worthy. 2018. “Assessing Government Transparency: An Interpretive Framework.” *Administration and Society* 50 (4): 501-26.

Miller, William R., John L. Sorensen, Jeffrey A. Selzer, and Gregory S. Brigham. 2006. "Disseminating Evidence-Based Practices in Substance Abuse Treatment." *Journal of Substance Abuse Treatment* 31 (1): 25-39.

Moore, Mark. 1995. *Creating Public Value: Strategic Management in Government*. Cambridge, Mass: Harvard University Press.

Moore, Mark. 2013. *Recognizing Public Value*. Cambridge, Mass: Harvard University Press.

Mulgan, R. 2014. *Making open government work*. Macmillan International Higher Education.

Nielsen, Steffen B., Rudi Turksema, and Peter van der Knaap. Eds. 2015. *Success in Evaluation: Focusing on the Positives*. New Brunswick: Transaction.

Nutley, Sandra M., Isabel Walter, and Huw TO Davies. 2007. *Using Evidence: How Research Can Inform Public Services*. Bristol: Policy Press..

Opperman, Kai, and Alexander Spencer Eds. 2016. "Fiascos in Public Policy and Foreign Policy." Special issue *Journal of European Public Policy* 23 (5): 643-787.

Page, Stephen, Melissa Stone, John Bryson, and Barbara Crosby. 2015. "Public Value Creation by Cross-Sector Collaborations: A Framework and Challenges of Assessment." *Public Administration* 93 (3): 715-732.

Patashnik, Eric. 2008. *Reforms at Risk*. Princeton: Princeton University Press.

Peters, Tom, and Robert Waterman. 1982. *In Search of Excellence*. New York: Harper Collins.

Rochlin, Gene. Ed. 1996. "New Directions in Reliable Organization Research." Special issue, *Journal of Contingencies and Crisis Management* 4 (2): 55-111.

Roe, Emery, and Paul Schulman. 2008. *High-Reliability Management*. Stanford: Stanford University Press.

Rosenbloom, David. 1983. "Public Administrative Theory and the Separation of Powers." *Public Administration Review* 43 (3): 219-227.

Ross, Michael L. (May 2015). "What Have We Learned about the Resource Curse?" *Annual Review of Political Science* 18 (2): 239–259.

Rothstein, Bo. 2012. "Political legitimacy for public." *The Sage handbook of public administration*, 407-419.

Sabel, Charles, and Jonathan Zeitlin. 2012. "Experimentalist Governance." In: David Levi-Faure Sabel Ed. *Oxford Handbook of Governance*. Oxford: Oxford University Press, DOI: 10.1093/oxfordhb/9780199560530.013.0012

Schon, Donald. 1971. *Beyond the Stable State*. New York: Random House.

Sørensen, Eva. 2017. "Political Innovations: Innovations in Political Institutions, Processes and Outputs." *Public Management Review* 19 (1): 1-19.

Suchman, Mark C. 1995. "Managing Legitimacy: Strategic and Institutional Approaches." *Academy of Management Review* 20 (3): 571-610.

Tamanaha, Brian. 2007. *A Concise Guide to the Rule of Law*. Legal Studies Research Paper Series #07-0082. St John's University School of Law. http://content.csbs.utah.edu/~dlevin/conlaw/tamanaha-rule-of-law.pdf

Tyler, Tom R. 2001. "A psychological perspective on the legitimacy of institutions and authorities." In J. Jost, & B. Major (Eds.), *The psychology of legitimacy*. Cambridge: Cambridge University Press.

Van Assche, Kristoff, Raoul Beunen, Martijn Duineveld. 2011. "Performing Success and Failure in Governance: Dutch Planning Experiences." *Public Administration* 90 (3): 567-81.

Van der Heijden, Jeroen. 2011. "Institutional Layering: A Review of the Use of the Concept." *Politics* 31 (1): 9-18.

Van der Wal, Zeger. 2017. *The 21st Century Public Manager*. London: Palgrave.

Waldron, Jeremy 2016. The Rule of Law. In E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy,* https://plato.stanford.edu/archives/fall2016/entries/rule-of-law/.

Walker, Brian, David Salt. 2006. *Resilience Thinking: Sustaining Ecosystems and People in a Changing World*. Washington: Island Press.

Weick, Karl, and Kathleen Sutcliffe. 2011. *Managing the Unexpected*: *Resilient performance in an age of uncertainty*. New York: Wiley.

World Health Organization 2017. *Assessment Report of the Global Vaccine Action Plan*. https://www.who.int/immunization/web_2017_sage_gvap_assessment_report_en.pdf

Internet sources

www.skybrary.aero/index.php/Aviation_Safety_Performance_Reports_and_Statistics (consulted 8 February 2018)

Figures and tables

Figure 1. The successful public governance assessment funnel

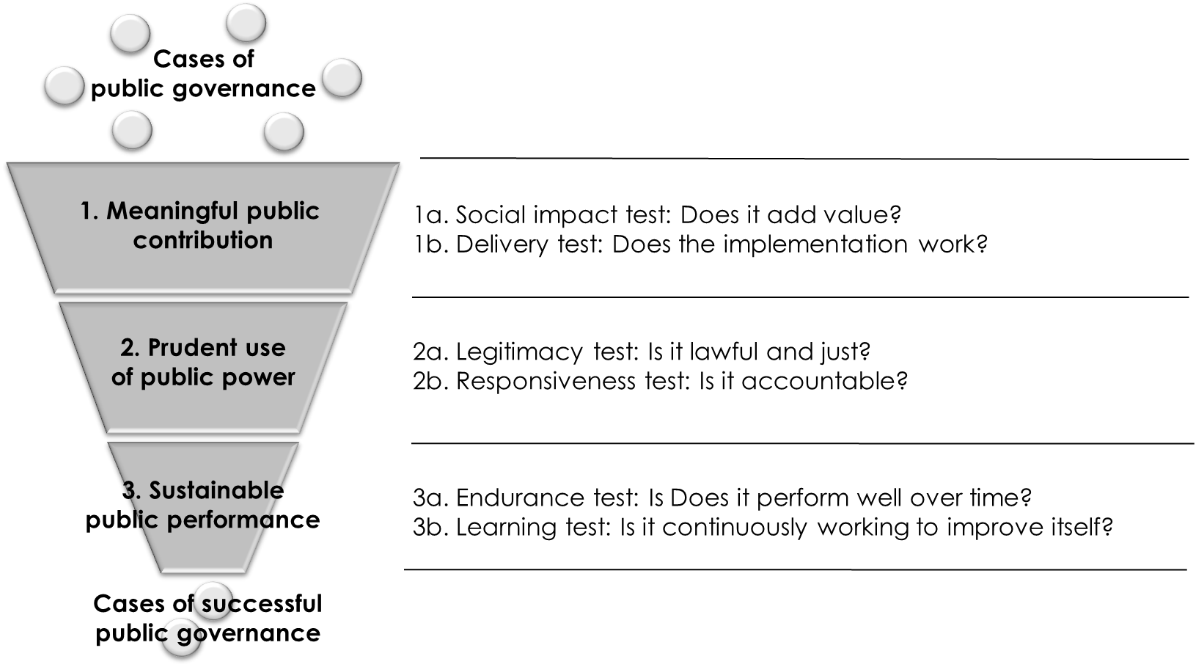

Table 1. The assessment funnel: operational framework for case assessment

Criteria	**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?

Table 2. Applying the assessment funnel

Case reviewer	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
Criteria	**GI Bill**			**Future Petroleum Fund**			**HIV/Aids**		
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, -, --, ---.

Figure 2. An overview of the case assessments