

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

Abstract

Journalists, parliaments, watchdog institutions, and public administration scholars devote considerable energy to targeting and dissecting failures in government. Studies and case-studies of policy, organizational and institutional failures in the public sector have figured prominently in public administration curriculums and classrooms. Such a focus on government failures provides students with cautionary tales and theoretical tools for understanding how things can go badly wrong in public policymaking and service delivery. Considerably less attention has been paid to instances of governments performing well. To address this imbalance, this article proposes an assessment framework designed to allow students to systematically identify, comprehensively assess and carefully interpret instances of successful public governance. Set up as a funnel consisting of three sequential assessment filters, each containing specific tests, the framework helps students operationalize 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. We discuss how the framework can be used in classroom settings to generate systematic consideration of key challenges of assessing and learning from success in government.

Keywords: policy success, high-performing organizations, evaluation, case teaching, positive public administration

Learning to Analyze Government Successes

The failures and faults of government can cause immense damage to societies at large and the lives of individual citizens. Journalists, parliaments, watchdog institutions, and public administration scholars therefore devote considerable energy to targeting and dissecting failures in government. These efforts gave rise to a vast array of insights in ‘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 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.

Unfortunately, our current public discourses and 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). 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, there appears to be a certain coyness about similarly applying a success prism (high performance, high legitimacy, high robustness) to the public sector. 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, 2021; 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.

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); and successful collaborative governance and network management (Dickinson and Sullivan 2014; Page et al. 2015; Cristofoli , Meneguzzo, and Riccucci 2017). Similarly, we integrate research on successful 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).

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. The following sections first outline the framework, and then provide a step by step approach for its application in the classroom.

The Funnel Framework

Our focus lies with offering a discursive tool for assessing the nature and degree of success of '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 policies, projects, programs, service delivery mechanisms, organizations and collaborations. Further, we assume governance success cannot just 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 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.

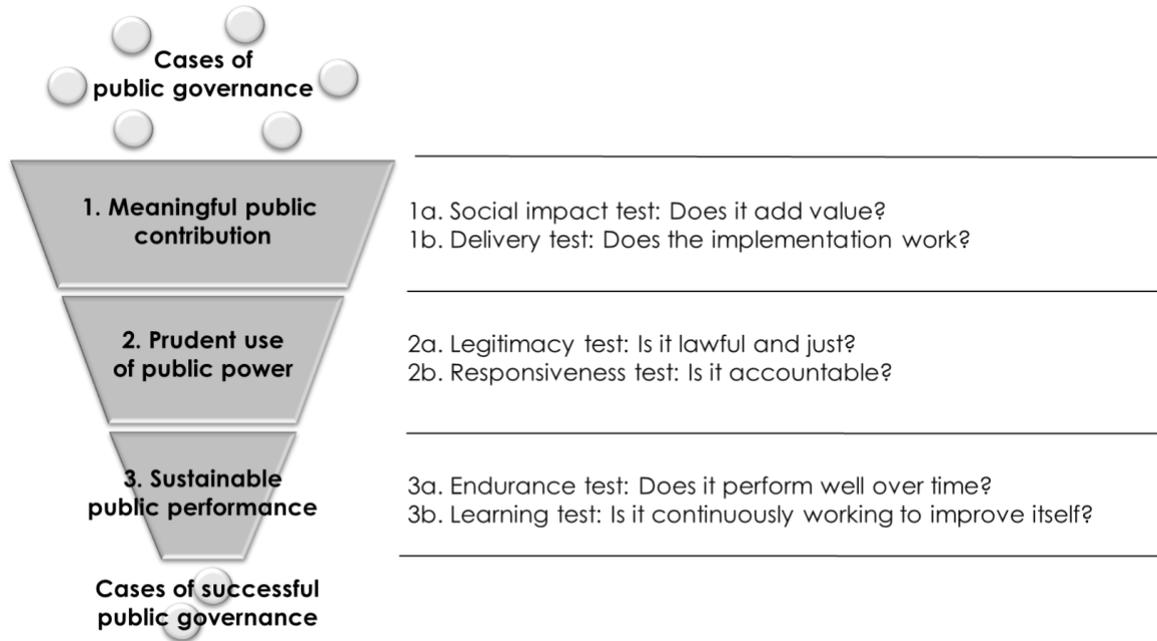
These three criteria are not to be thought of as a multi-perspectivist framework of different constitutional powers (law, politics and management, Rosenbloom 1983), concurrent or competing values (Hood 1991; Lindquist and Marcy 2016) nor as a normative hierarchy (Fischer 1995), though they derive from underlying normative traditions (Bovens et al. 2008; Meijer et al., 2018). As we will elaborate below, they are constructed as a *three-stage funnel* with each stage comprising a set of evaluative tests. Only cases that pass all three stages of the assessment funnel process with flying colours can be considered to be ‘complete successes’, whereas cases that do not even make it past the first sets of tests can be considered a failure.

In terms of 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 criteria) or conflicted (persistent disagreement among different evaluators about what scores should apply on certain criteria) success). While the funnel framework can be a useful evaluation tool for both high and low performing cases, our primary interest in offering this tool is its potential for identifying and analyzing cases on the high end of the success spectrum.

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

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 governance initiative 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 exercising power or limiting

liberties, 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. 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 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. Immaterial costs and benefits can only be expressed in crude estimates. Moreover, 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. Hence, what results is necessarily and ultimately a qualitative judgement.

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 and execution 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, 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 (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 behaviours of different target populations (Howlett, Mukherjee, and Woo 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. This can vary per context. 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.

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. Hungarian laws to curtail judicial independence and freedom of expression, for example, are declared in conflict with the Rule of Law by the European Parliament.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 et al. 2011). Multiple ‘publics’ with heterogeneous and potentially

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>

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 (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 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 preoccupation 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' (Schon 1971; Goodsell 2011).

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, 2016; Howlett, Capano and Ramesh, 2018; Ansell, Sorensen and Torfing, 2020). How well they adapt over time is crucial to the endurance of success (Compton and 't Hart 2019; Luetjens et al. 2019). Patashnik's (2008) study about what explains the survival of some and the demise of other general interest reforms evolves 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 (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 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 institutionalisation of a learning culture in the global civil aviation safety regime (www.skybrary.aero 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. <i>The social impact test:</i> Does it add value?	<ul style="list-style-type: none"> • 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. <i>The delivery test:</i> Does implementation work?	<ul style="list-style-type: none"> • 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. <i>The legitimacy test:</i> Is it lawful and just?	<ul style="list-style-type: none"> • Are governance processes and outcomes in accordance with the Rule of Law and perceived as just and fair by all stakeholders?
2b. <i>The responsiveness test:</i> Is it accountable?	<ul style="list-style-type: none"> • 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. <i>The robustness test</i> : Does it perform well over time?	<ul style="list-style-type: none"> • Are considerations of long-term sustainability given due attention in the institutional design and management of the initiative?
3b. <i>The learning test</i> : Is it continuously working to improve itself?	<ul style="list-style-type: none"> • 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 have integrated the use of the framework in a number of both undergraduate, postgraduate, and executive courses on successful public governance. From these experiences, we developed the following teaching approach.

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 make 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:

- The Public Impact Observatory produced by the Center for Public Impact, a not-for-profit offshoot of the Boston Consulting Group:
<https://www.centreforpublicimpact.org/insights>
- The Collaborative Governance Database, curated by a global network of collaborative governance scholars:
<https://collaborativegovernancedatabase.sites.uu.nl/>
- 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>
- The cases included in several 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; Luetjens et al, 2019; DeLaPorte et al., forthcoming; Howlett et al., forthcoming).

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, 2014). 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 programmes. 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.’

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. <i>The learning test</i> : Is it continuously working to improve itself?	+ +++ ++	++ ++ ++	++ ++ ++

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

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 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.

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 (Majone, 1989).

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 judgements.

Finally, students can be encouraged to follow in the footsteps of, e.g. 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).

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 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). Great policy successes in other words can serve as game-changing precedent-setters, altering for the better what we know governments can deliver if wisdom, courage, astuteness and coincidence converge around a social cause.

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-Coordination*. 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, Paul 't Hart and B. Guy Peters. Eds. 2001. *Success and Failure in Public Governance: A Comparative Analysis*. Cheltenham: Edward Elgar.
- 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.
- Capano, Giliberto and Jun-Jie Woo (2017). "Resilience and Robustness in Policy Design: A Critical Appraisal." *Policy Sciences* 50(1): 1–28
- Carpenter, Daniel. 2001. *The Forging of Bureaucratic Autonomy: Reputations, Networks and Policy Innovation in Executive Agencies, 1862-1928*. Princeton: Princeton University Press.
- Coglianesse, Cary. 2016. *Achieving Regulatory Excellence*. 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.

Connolly, J., Reid, G., & 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, 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.

DelaPorte, Caroline, Guony Eydal, Jaakko Kauko, Daniel Nohrstedt, Bent-Sofus Tranøy, Paul 't Hart. Eds. (forthcoming). *Successful Public Policy: Lessons From the Nordic Countries*. Oxford: Oxford University Press.

Deutsch, Karl. 1960. *The Nerves of Government*. New York: Free Press.

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 Bøgh 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.
- Herek Gregory, Irving L. Janis, Paul Huth. 1987. "Decision Making During International Crises: Is Quality of Process Related to Outcome?" *Journal of Conflict Resolution*.31(2):203-226.
- 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.
- Howlett, Michael, Giliberto Capano & M. Ramesh. 2018. "Designing for Robustness: Surprise, Agility and Improvisation in Policy Design." *Policy and Society* 37 (4): 405-421.
- Howlett, Michael, Evert M. Lindquist, Grace Skogstad, Geneviève Tellier and Paul 't Hart. Eds. (forthcoming). *Successful Public Policy: Lessons From Canada*. Oxford: Oxford University Press.
- 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.

Luetjens, J., & 't Hart, P. (2019). *Governing by looking back: learning from successes and failures*. Australian Public Service Review Panel.

Luhman, Niklas. 1995. *Social Systems*. Stanford: Stanford University Press.

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

Majone, Giandomenico. 1989. *Evidence, Argument and Persuasion in the Policy Process*. New Haven: Yale University Press.

Majone, Giandomenico. 1996. Temporal Consistency and Policy Credibility: Why Democracies Need Non-Majoritarian Institutions. Florence: European University Institute, Working Paper RSC No 96/57.

McConnell, Allan. 2010. "Success, Failure, or Somewhere In-Between?"

McDonald III, B. D. (2021). Teaching in uncertain times: The future of public administration education. *Teaching Public Administration* 39 (1): 3-8.

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.
- Mushkat, M. (2001). Teaching Case Analysis to Public Managers in Academic Settings. *Teaching Public Administration*, 21(1), 37-45.
- 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.
- Schafer, Mark and Scott Crichlow. 2010. *Groupthink vs High Quality Decision Making in International Relations*. New York: Columbia University Press.

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. 1990. *Why People Obey the Law*. New Have: Yale University Press.

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 source

[www.skybrary.aero/index.php/Aviation Safety Performance Reports and Statistics](http://www.skybrary.aero/index.php/Aviation_Safety_Performance_Reports_and_Statistics)
(consulted 8 February 2018)