Designing for Policy Success

Mallory E. Compton, Joannah Luetjens and Paul 't Hart

Designing for Policy Success

Mallory E. Compton
Texas A&M University (USA)

Joannah Luetjens
Utrecht University (Netherlands)

Paul 't Hart
Utrecht University (Netherlands)

Abstract

Amidst the general mood of skepticism about the problem-solving capacity of governments in the face of 'wicked problems', it is easy to overlook that at times governments do manage to design and implement public policies and programs quite successfully. In this paper, we build on an emerging area of 'positive evaluation' research into public policy successes (Bovens et al 2001; McConnell 2010; Nielsen et al 2015). Using the conceptual tools emanating from this research and drawing on a corpus of 33 such cases (Compton and 't Hart 2019; Luetjens et al, 2019), we draw inferences about the contexts, strategies, and practices that are conducive to policy success. We find compelling evidence that process inclusivity is a pivotal factor, but certainly not the only one, on the path to policy success. Variation in the degree of innovation and the pace of change also emerge as interdependent and important factors.

Keywords

policy success, qualitative comparative analysis, performance, positive evaluation

Introduction

The policy analysis field is rediscovering its long-held ambition (Bobrow and Dryzek, 1987) to become a design science. New handbooks (Howlett and Mukherjee, 2018), monographs (Peters, 2018), textbooks (Howlett, 2019), compact surveys of the field (Peters et al, 2018), and even a dedicated journal (*Policy Design and Practice*), have emerged in this pursuit. The goal of these endeavors is to discover strategies and practices for effective policy design. To this end, policy design scholars analyze policy instrument choices and policy implementation processes, and they identify the mechanisms which link design features to implementation processes and policy outcomes (Busetti and Dente, 2018; Capano et al, 2019). To translate these findings into design maxims and practices to significantly increase the odds of policy success is the ultimate ambition of the endeavor.

This (re)turn to policy analysis as design holds considerable promise to give the field a sense of purpose and application-oriented allure. Such direction has been lacking since the field lost ground to an emphasis on developing general descriptive and explanatory models of policy processes and implementation (Linder and Peters 1987, 1988). Likewise, the design-oriented, hands-on, and applied agenda of the policy sciences has been overcome by a recognition of emerging megatrends such as globalization and the turn from government to governance on public policymaking (Howlett, 2014). Despite heroic efforts by some of their chief advocates (see Weible and Cairney, 2018), followers of this approach still have a long way to go before they have tangible, practical, and actionable strategies to offer to public policymakers.

Rekindling the spirit of design within the policy sciences is, thus, a good and necessary enterprise. At the same time, however, revivalists of design-oriented policy analysis have demonstrated a curious tendency: they have not devoted much attention to the kind of policy outcomes towards which they ultimately strive – presumably 'valuable' or indeed 'successful' public policies. Most have stuck with abstract notions of policy effectiveness, such as 'creating a frame for action that may shape a range of policy responses' (Peters et al, 2018: 19-20). While conceptually elegant, such formulations are difficult to operationalize and do not allow analysts to address the well-documented conceptual and methodological challenges inherent in making goal achievement the main criterion for policy success (Bovens et al, 2006; McConnell, 2010; Howlett, 2012). The policy design movement needs a richer and more relevant conceptualization of its key dependent variable to serve as a rock-solid foundation on which to pursue the ambition of identifying and testing the impact of various design principles and policy instrument mixes on policy success.

If we are serious about reviving policy analysis as a design enterprise, and if we agree that such an enterprise should be empirically rooted in careful study of 'what works' (and what fails to work), then we must embrace the challenge of developing methods to assess the success or otherwise of public policies. A fundamental challenge in this task will be to better represent the complex, dynamic, and contested nature of public policymaking in our conceptualization and measurement of the outcome of interest – successful policies and programs. Furthermore, we must do the empirical legwork of applying those standards to cases within and across policy sectors and jurisdictions (e.g. Bovens et al, 2001; Grace et al, 2017; De Francesco and Maggetti, 2018) and then identify the patterns of conditions, factors, and mechanisms common to cases of policy success. *Only then* may we say that we have established 'what works' in successful public policies and programs.

In this article, we provide the tools for doing so, and we explore what this may look like in

comparative case analysis. Building upon foundational efforts by Bovens et al (2001) and Mc-Connell (2010) we first present an assessment tool for identifying dimensions and degrees of policy success. We then introduce three design factors whose (combined) impact on different dimensions of policy success we examine empirically: process inclusivity, degree of innovation, and the pace of change. We continue by presenting the design of this approach and reporting on the findings of a fuzzy-set Qualitative Comparative Analysis (QCA) covering 33 case studies taken from two recent collaborative research projects on policy success (Compton and 't Hart, 2019 and Luetjens et al, 2019 - brief case descriptions can be found in the Appendix). We conclude this article by identifying the promises and pitfalls of our approach for the effort to rebuild policy analysis as a design science.

Identifying policy successes: an assessment tool

Policy successes and policy failures are construed in stories. Undoubtedly 'events' – real impacts on real people – are important factors in the shaping of those stories, but their importance is neither given nor straightforward. To claim that a public policy, program, or project 'X' is a 'success' is effectively an act of interpretation (Bovens et al, 2001; Kay and Boxall, 2015). These interpretations are informed by the evaluative and oversight work of professional bodies and think tanks (think of the OECD's rankings or the PISA scores), government regulators, auditors and evaluators, and ad-hoc reviews and inquiries. It amounts to giving a strong vote of confidence to certain acts and practices of governance. It singles them out, elevates them, validates them. Conversely, labeling a particular instance or episode as an outright failure entails sharp negative feedback, a repudiation of a now compromised past. It sets in motion institutional rituals in which account-giving is demanded, blame is apportioned, and learning must be shown to take place so as to avoid repetition (Van Berkel et al, 2016; Olavarria-Gambi, 2018).

For such an act to be consequential it needs to stick: others must be attracted to its appeal and they need to emulate it. The claim 'X is a success' must become a widely accepted and shared narrative. When it does, 'success' becomes performative: X looks better and better because so many say so, so often (Van Assche et al, 2012). When the narrative endures, X becomes enshrined in society's collective memory through repeated re-telling and other rituals. Examples of the latter include the conferral of awards on people or organizations associated with X, who are subsequently invited before captive audiences to spread the word; the high place that X occupies in rankings; the favorable judgments of X by official arbiters of public value in a society, such as audit agencies or watchdog bodies, not to mention the court of public opinion.

Trying to establish 'what works' is treacherous terrain, and we must tread carefully (see Howlett, 2014) to adopt a transparent and widely applicable conceptualization of 'policy success' to employ in the project at hand. Based on this conceptualization of 'success,' we must then identify a set of research tools to identify and characterize those 'successes.' To get there, we posit that policy assessment is necessarily a multi-dimensional, multi-perspectivist, and political process. At the most basic level we distinguish between two dimensions of assessment. First, the programmatic *performance* of a policy: success is essentially about designing smart programs that will really have an impact on the issues they are supposed to tackle, while delivering those programs in a manner to produce social outcomes that are valuable. There is also the political *legitimacy* of a policy: success is the extent to which both the social outcomes of policy interventions, and also the manner in which they are achieved, narrated and widely accepted as meaningful and appropriate by stakeholders and accountability forums (Fischer 1995; Hough et al 2010).

The relation between these two dimensions of policy evaluation is not straightforward. There can be (and often are) asymmetries: politically popular policies are not necessarily programmatically effective or efficient, and vice versa. Moreover, there is rarely a single shared normative and informational basis upon which all actors in governance processes assess performance, legitimacy, and endurance (Bovens et al. 2001). Many factors influence beliefs and practices through which people form judgments about governance.

Heterogeneous stakeholders hold diverse vantage points, values, and interests regarding a policy, and thus may experience and assess it differently. An appeal to 'the facts' does not necessarily help settle these differences. Like policymaking, policy evaluation occurs within a context of multiple, often competing, cultural and political frames and narratives, each privileging different facts or considerations. As contexts change, new events occur, or players enter and leave the arena over time, the appeal and momentum of alternative frames and narratives can shift. Evaluation is inherently political in both approach and implications, no matter how deeply espoused the commitment to scientific rigor of many of its practitioners. This is not something we can get around. It is something we have to acknowledge without sliding into thinking that it is *all and only* political, and that therefore 'anything goes' when it comes assessing the success or otherwise of a policy (Bovens et al. 2006).

We build upon prior work in developing a useful model for evaluation. Adding to Bovens and 't Hart's programmatic-political dichotomy, McConnell (2010) added a third perspective to produce a three-dimensional assessment map. We have adapted this three-dimensional assessment for our purposes (see also Newman, 2014) and added an additional – *temporal* – dimension. It thus consists of the following components:

Programmatic assessment – This dimension reflects the focus of 'classic' evaluation research on policy goals, the theory of change underpinning it and the selection of the policy instruments it deploys – all culminating in judgments about the degree to which a policy achieves valuable social impacts.

Process assessment - The focus here is on how policy design and decision-making are organized and managed, and whether these processes contribute to vigilant public problem-solving in ways that enhance policy effectiveness through the application of rigorous deliberative processes in which evidence, argument and persuasion are given a wide berth (Majone, 1989).

Political assessment – This dimension assesses the degree to which policymakers and agencies involved in driving and delivering the policy can build and maintain supportive political coalitions, and the degree to which policymakers' association with the policy enhances their reputations. In other words, it examines both the political requirements for policy success and the distribution of political costs/benefits among the actors involved.

Endurance assessment - The fourth dimension adds a temporal perspective. We surmise that the success or otherwise of a public policy, program or project should be assessed not through a one-off snapshot but as a multi-shot sequence or episodic film ascertaining how its performance and legitimacy develop over time. Contexts change, unintended consequences emerge, surprises are thrown at history: robustly successful policies are those that respond to these dynamics through institutional learning and flexible adaptation in program (re)design and delivery, and through political astuteness in safeguarding supporting coalitions and maintaining public reputation and legitimacy.

Taking these four dimensions into account, we propose the following definition of a policy success:

> *A policy is successful to the extent that it purposefully creates widely valued social outcomes through rigorous processes and manages to sustain this performance for a considerable period of time, even in the face of changing circumstances.*

Table 1 presents an assessment framework that integrates these building blocks. Articulating specific elements of each dimension of success – programmatic, process, political, endurance – in unambiguous and conceptually distinct terms, this framework lends a structure to both contemporaneous evaluation and dynamic consideration of policy developments over time.

Table 1. A Policy Success Assessment Map

Programmatic assessment: **Purposeful and valued action**	**Process assessment:** **Thoughtful and fair policymaking practices**	**Political assessment:** **Stakeholder and public legitimacy for the policy**
A well-developed and empirically feasible *public value proposition* and *theory* of change underpins the policy.	The policy process allows for *rigorous deliberation* about the relevant values and interests; the hierarchy of goals and objectives; contextual constraints; the (mix of) policy instruments; and the institutional arrangements and capacities necessary for effective policy implementation.	A relatively broad and deep political *coalition* supports the policy's value proposition, instruments and current results.
Achievement of (or considerable momentum towards) the policy's intended and/or of other *beneficial social outcomes.*		Association with the policy *enhances the political capital* of the responsible policy-makers.
Costs/benefits associated with the policy are *distributed equitably in society.*	Decision-making processes incorporated balanced consideration of a wide range of evidence, expertise and advice.	*Association with the policy enhances the organizational reputation* of the relevant public agencies.
Temporal assessment		
Endurance of the *policy's value proposition* (the proposed 'high-level' ends-means relationships underpinning its rationale and design, combined with the flexible adaptation of its 'on-the-ground' and 'programmatic' features to changing circumstances and in relation to performance feedback). Degree to which the policy's programmatic, process, and political *performance is maintained* over time Degree to which the policy confers *legitimacy on* the broader political system		

Source: Compton and 't Hart (2019) and Luetjens et al (2019)

How policy successes happen: exploring three design factors

Now that the outcome of interest – policy success – has been conceptualized, we move towards presenting the three design factors whose (singular and combined) contribution to the achievement of policy success we want to examine. The policy design literature offers many potentially relevant candidates, but the nature of our research design – fuzzy set analysis, which will be explained below – limits the number of conditions that we can consider. We have selected the following three, each of which has been the subject of sustained debate in the field (though many others are listed in extensive overviews such as Hood and Margetts, 2007):

Process inclusivity: open and cooperative?

Theories of governance across the social sciences often consider the importance of including or representing diverse interests in decision-making. Much of this work has focused on the promise of broadly inclusive, consultative, or collaborative decision-making in the production of legitimate, just, creative, and sustainable solutions to public problems (Torfing, 2019). There are multiple theoretical foundations upon which such associations may be based.

First, work on collective action problems has shown how self-organized bottom-up cooperation among directly affected stakeholders (citizens) may be comparatively advantaged in producing sustainable institutions (Ostrom, 1990). Local knowledge of contexts and informal institutions may lead to better tailored policy solutions, and collaboration in the process may facilitate greater trust and legitimacy in the process and outcome. Second, public management theory has long reiterated the importance of including actors who are intimately involved in front-line implementation and public service delivery at the 'front end' of policy design. These can include not only street-level bureaucrats in line agencies but also citizens and civil society groups who effectively co-produce public services (Alford, 2009) as well as non-profits and firms contracted to deliver public services. Including them in policy design enables their experiential knowledge of 'what *actually* works' (and what definitely does not) to weigh into decisions about the (mix of) policy instruments to be deployed and the way in which implementation processes are being structured (Gouillart and Hallett, 2015). The inclusion of and deliberation with a range of societal actors in forums previously reserved for elected officials is valued not only on the basis of democratic principles of equity and representation, but also for practical gains from more effective and legitimate policy design and public value production (Bryson et al, 2014).

We therefore expect that the level of inclusivity of policy design processes may be an important condition of a public policy, program, or plan. Tightening or widening the circle of stakeholders, merely informing or actively involving them, and at what stage in the policy design process, are key options around inclusivity that policy designers and network managers can purposefully choose (Klijn and Koppenjan, 2016: 252). But does it matter if they do? In our empirical study we shall examine whether and how process inclusivity conduces towards policy success.

Degree of innovation: exploitation or exploration?

Policy designers and analysts typically face two alternatives when it comes to making design decisions. The first option is to refine and extend existing policies, programs, and instruments by drawing from experience both within and outside their jurisdiction. To improve and to learn, policymakers are encouraged to cast their attention to instances of success to identify and then exploit 'what works' in their own professional domains (Bennett, 1991). Acclaimed 'best practices' from other professions and jurisdictions both at home and abroad can also serve as

models. In its ideal typical form, this process of policy transfer is expected to follow a certain sequence. First, policymakers identify a policy or program area within their own jurisdiction that they want to make progress on. Solutions are then adopted that copy past experiences of their own or programs that have run abroad. Exploitation thus involves continuity through inheritance, the use of analogies and imitation. However, when the keenness to consolidate and copy trumps the discipline to first experiment and adapt in the new context, exploitation-based policy designs may end in disappointment.

The second possibility is to pursue policy options that are new and thus untested within the jurisdiction, sector, or network at hand. This explorative mode involves at a minimum the explicit tailoring to local conditions of programs and policy instrument mixes borrowed elsewhere, and more radically the development of new frames on existing conditions and problems, the engagement of new stakeholders or a realignment of existing stakeholders, a search for knowledge that departs from established truths, the discovery of new values or the recalibration of existing values and calculi (Torfing and Triantafillou, 2016). In combination, these may generate completely new policy designs and delivery mechanisms that have the potential to increase public problem-solving capacities and forge progress on issues that defied existing governance repertoires.

There is no 'right' choice between these two modes as they each have their own risks and rewards. Established practices that over time have been tried, tested, and improved on through learning processes minimize the risk of surprise. Policy design processes relying exclusively on established processes, however, run the risk of creating rigid approaches which prove difficult to adapt once circumstances change (Hacker, 2004). Conversely, the potential promise of new possibilities and practices is enticing but inherently laden with uncertainty and thus risk. Strong reliance on exploration can mean that undeveloped ideas turn into costly investments which may not yield the expected public value (March, 1991).

In our research design, we include the condition of 'degree of innovation' to ascertain the extent to which key elements of policy design were imitated or invented. Invention is taken to mean the development of something that is entirely new to the jurisdiction or sector in which it is being applied, though it may not necessarily be new to the rest of the world (Rogers, 2003: 43). Imitation, in contrast, suggests that core elements of the policy's design were inherited, borrowed, transferred, or emulated.

Pacing of change: steps or leaps?

Much policy design activity is remedial and aspirational at the same time. It rests on what has famously been called 'problemistic search' (Cyert and March, 1963; Posen et al, 2018), aiming to move issues, groups, or communities away from a status quo that has undesirable features towards a more agreeable future state. A key question facing policy designers is how such movement can be accomplished. This question has complex behavioral (how empirically valid is the 'theory of change' upon which the current policy design and in particular its policy instrument mix relies?), institutional (what changes in existing capacities, rules, repertoires and routines of implementing actors are required to enable effective delivery of the policy?), and political (how to build and maintain sufficiently powerful coalitions that support the social change sought by the policy?) ramifications.

This is delicate work: the status quo in any social system is a negotiated order, and efforts to purposefully change it will bring discomfort to all or parts of that system, expose patterns of privilege and disadvantage, question institutionalized beliefs and norms, and generally require

people to 'do things differently'. Change is fraught with uncertainty and conflict, and systems work very hard – consciously and unconsciously – to avoid it. Policy designers whose reach (the scope of their ambitions) exceeds their grasp (their ability to overcome or circumnavigate inertia and pushback) will be punished by seeing their best laid plans come to naught, or worse.

In dealing with these challenges, one of the key conditions that policymakers can manipulate involves what Heifetz et al (2009) call 'pacing' the work of change: assessing when to push the accelerator, when to switch gears, and when to hit the brakes. In the policy sciences this has been the subject of robust debate. On the one hand, there are those who argue that within polycentric systems, purposeful social change can only come about through relatively small steps: a succession of compromises that work in roughly the same direction (Lindblom, 1979; Rothmayr Allison and Saint-Martin, 2011; Hayes, 2017). On the other hand, however, are those who argue that incremental compromises are simply not good enough when addressing large and urgent challenges. A status quo may be simply unacceptable, or even threatening. To effect sufficient change, it may be necessary to muster ambition, vision, power, and capacity to move more boldly and quickly than 'politics as usual' tends to allow (Dror, 1986, 2001). Or, change may require clever use of 'windows of opportunity' that – sometimes predictably but more often erratically – present themselves in the form of incidents, outrages, and crises (Boin et al, 2009; Hogan and Feeney, 2012).

The condition 'pace of policy adoption' has been incorporated into our design to address this debate. We examine to what extent slow-paced and incrementally constructed policy, as opposed to the forging of fast-paced 'crash or crash through' (the dictum of Australia's most reforming prime minister Gough Whitlam, 1972-74) policy change, are conducive to policy success.

A comparative examination of thirty-three cases: research design

Policy processes involve many factors which operate at multiple levels and in combination with one another to shape outcomes of interest. The study of policy success thus encourages an explanation using combinatory logic. In this analysis, we are not seeking causal explanation or formal comparison, nor do we endeavor to arrive at universal (or even external) generalizability or estimation of average effects, let alone aim to identify (probabilistic) empirical regularities. Our goal is to study how policy outcomes (dimensions and levels of policy success) are produced through the confluence and interplay of certain 'designable' features of the policies in question. Fuzzy set Qualitative Comparative Analysis (fsQCA) is well-suited to this task.

QCA (Schneider and Wagemann, 2012) is a comparative case-oriented research approach suitable for unearthing complex patterns of causality across a universe of cases. It entails a collection of techniques based on set theory and Boolean algebra, enabling logical reasoning about actual cases, their conditions, and how outcomes emerge from a combination of these conditions. By comparing configurations and pooling similar cases together, it allows researchers to explore similarities and differences across comparable cases. QCA has both descriptive and explanatory uses, which may include summarizing data, creating typologies, evaluating existing hypotheses, and developing new theories.

The main difference between QCA and other more quantitative research methods lies in the idea of causality underpinning the approach (Ragin, 2008; Rihoux and Ragin, 2009; Schneider and Wagemann, 2012). Methods such as statistical analysis tend to imply mono-causality and focus on the estimate of each independent variable's separate effect on the variation of the dependent variable. In contrast, QCA aims to produce multi-causal explanations. It focuses on combinations of conditions rather than single variables and does not assume that a unique 'so-

lution' can account for the occurrence and non-occurrence of a particular outcome (Vis, 2012).

In QCA methodology, outcome values within and across cases are accounted for in terms of distinct configurations of conditions. Once cases have been selected, QCA-analysis proceeds by specifying the outcome dimensions that are the focus of analysis and then by viewing each of these dimensions as a 'set' in which cases can have varying degrees of membership (Schneider and Wagemann, 2010). Here we employ fuzzy-set QCA (fsQCA) only as a data analysis technique that aims to identify empirical patterns in the data. FsQCA was selected as it allows us to make fine-grained distinctions both between and within sets (Ragin, 2008). That is, it captures variation across cases both in degree and in kind.

In fsQCA, explanations are expressed in terms of necessity and sufficiency. First, we check for the presence of necessary conditions and then verify whether there was overlap between necessary conditions for the presence of policy success and its absence (Schneider and Wagemann, 2012). Following this, we turn to the analysis of sufficiency through the construction of truth tables. The truth table effectively synthesizes how many cases adhere to a certain pattern, and if they consistently show the same outcome. If a particular pattern is consistent, then these rows are used in a minimization procedure which produces a particular solution term. Here, we opted for the conservative solution as no logical remainders (no empty truth table rows) appeared. To arrive at the conservative solution, sensitivity and robustness tests were conducted. This involves adjusting the consistency threshold to detect the robustness of the resulting solutions. The general standard is considered to be 0.80 (Schneider and Wagemann, 2012), which was run first. If consistency is set below 0.8, then it becomes increasingly difficult to maintain that a relationship exists. From 0.80 to 0.83, the solution terms obtained were robust. Setting the threshold above 0.85 produced no solution terms. Setting the threshold at 0.83 allowed us to explain more of the cases.1

Case Selection

Given our goal of identifying the combination of conditions which are conducive to successful outcomes, we limited our sample to cases of success. Each case in our sample was drawn from two larger collaborative projects, which collectively produced 33 in-depth case studies of policy success[1]. Candidate cases were identified as a 'great policy success' by expert scholars in the relevant policy domain along more than one but preferably all of the four success dimensions distinguished above: procedural, programmatic, political and endurance assessment. We sought cases of seen successes, which are not only successful (which we might posit is a more common condition than is popularly acknowledged), but also recognized as such. To find these gems, i.e., the projects from which we drew our cases, we consulted with experts and academics in a range of policy domains (environmental, public works or infrastructure, social welfare, healthcare, technology, and economic policy) to identify cases meeting the criteria for 'policy success.'

By offering insight into occurrences of policy success across varied contexts, the case studies in the two collaborative projects were designed to increase awareness that government and public policy actually work remarkably well, at least some of the time. In selecting cases for inclusion in the two collaborative volumes, some cases were removed due to concerns or disagreements between experts on a case's level of success. The cases were deliberately chosen to cover a broad range of issues, challenges, and policy sectors. These include cases of different modes (from

---

1 — The consistency score indicates the extent to which the configuration is always associated with a given outcome. The coverage score reflects the percentage of cases that the configuration can explain. That is, how well a combination of conditions can adequately explain an outcome of interest..

top-down central steering to open, deliberative and collaborative processes) and levels (from urban to the global) of governance. Though somewhat skewed to countries consistently ranking among the best governed in the world, the volume includes cases of federal and unitary, parliamentary and presidential, and Westminster and consensual systems of government. All of the cases – except Brazil's *Bolsa Familia* scheme and healthcare performance in Singapore – come from countries within the OECD.

Restricting our sample to successful cases may concern those more accustomed to a more quantitative or probabilistic logic of inference. Though criteria for sample selection vary across the quantitative-qualitative divide (Mahoney and Goertz, 2006), it is agreed that 'the cases you choose affect the answers you get' (Geddes, 2003). Selecting cases on the value of the dependent variable can profoundly bias statistical findings, fouling generalization and average effect estimation (Heckman, 1976). There are, however, defensible reasons to violate the dependent variable rule and select only or mostly 'positive' cases (Brady and Collier, 2010). Case selection should be a deliberate and well-considered procedure tailored to the specific research question at hand and type of explanation sought (Brady and Collier 2010; King et al. 1994). In this project, we do not seek to estimate average effects. We instead seek to identify pathways to success.

Coding Protocol and Calibration

Translating information into membership scores for fuzzy sets (so-called 'calibration') requires clear concepts for conditions and the outcome. Our calibration strategy involved the development of a rubric or coding scheme to assign values for the outcome and the conditions. We opted for a four-value fuzzy scale. A four-value scheme is particularly advisable when researchers have a substantial amount of information about cases, but the nature of evidence is frequently not identical across cases. Similarly, with increasing levels of differentiation it becomes ever more difficult to identify both theory-based and empirically observable distinctions between the values. We accordingly first operationalized the four dimensions of policy success of Table 1 into a fuzzy set table, assigning four possible scores indicating 'membership of success' on each dimension (see Table 2). A case which scored 0.33 or less was considered out of the set, whereas a case that scored 0.67 or more was considered in the set.

Table 2. Operational coding rubric for dimensions and degrees of policy success

	Programmatic	*Process*	*Political*	*Endurance*
Fuzzy Set Value	*Valued impacts achieved*	*Rigor and vigilance in deliberation*	*Reputational impacts*	*Sustained performance*
Fully out of the set of complete policy successes (0)	*Limited* public value created but initiative *besieged by* major setbacks and/or negative unintended consequences	*Chronic* lack of vigilant practices of policy design, decision-making or implementation. Role of quality evidence in the policy process *non-existent* or *deeply contested*.	The policy has become a *heavy reputational burden* for its architects and supporters, posing an existential threat	Both goals and instruments of the policy are *prematurely terminated, superseded or very significantly changed*.

More out than in of the set of complete policy successes (0.33)	*Some* public value created but initiative *accompanied by* considerable setbacks and/or negative unin- tended consequences	*Sporadic but unsystematic* vigilance in policy design, decision-making and implementation practices. *Limited* consensus about and/or use of quality evidence in the policy process.	The policy has become a *reputational liability* for its architects and supporters, regularly taking up their time and/ or other resources to defend	Important components of goals and/or instruments of the policy *continue to be challenged* following their initial adoption
More in than out of the set of complete policy successes (0.67)	*Considerable* public value created but *punctuated* by occasional setbacks and/or negative unin- tended consequences	*Frequent but not universal* and institutionalized vigilance in policy design, decision-making and implementation practices. *Considerable* appetite for use of quality evidence in the policy process though its actual role less prominent than hoped or claimed by policymakers.	The policy enjoys broad legitimacy, with the *potential to confer reputational benefits* on its architects and supporters	The goals and/or instruments are *generally stable* but remain *open to change* through external events and/or recalibration
Fully in the set of complete policy successes (1)	*Much* public value created, virtually *unrestrained* by setbacks or nega- tive unintended consequences	*Continuously* vigilant policy design, decision-making and implementation practices. *Institutionally embedded* use of quality evidence in the policy process	Association with the policy *enhances the reputation* and *political legitimacy* of both its architects and its supporters	Policymakers are *capable of adapting* the goals and/or instruments of the policy to *anticipate and reflect* changing circumstances

Source: The Authors

We then operationalized the three conditions we expect to be associated with successful policy outcomes (see Table 3). The 33 cases were coded according to the rubric (see Appendix 2). Initially, each author was responsible for coding a third of the cases. To ensure the reliability of these initial coding decisions, each author coded an additional set of cases that had already been coded by a different author. These coding decisions were then compared to identify discrepancies. Some minor discrepancies existed but these concerned differences in degree rather than in kind. That is, no discrepancies were found as to whether the case should be considered 'in' the set (either fully, or more in than out). Similarly, there were no discrepancies as to whether a case should be considered 'out' (either fully, or more out than in). Any remaining points of doubt were resolved by an additional round of coding by each author and cross-checking one another's codes and justifications.

Two difficulties emerged in this process, which we dealt with in turn. First, QCA methodology requires each case to be scored on *all* assessment dimensions in order to be included in analysis. For this reason, we were forced to drop two cases from the sample: Avoiding the Global Financial Crisis and Nuclear Free New Zealand. The financial regulation case could not be scored on 'endurance' as the case focuses entirely on a single moment in time (the late 2008-early 2009 panic on financial markets), and New Zealand's non-nuclear stance could not be scored on 'process' as it was principally a one-shot symbolic policy by pronouncement. A second difficulty emerged in scoring the case of the Montreal Protocol to protect the ozone layer. As the only transnational case in the sample, we consulted a leading expert of international regimes and treaties to suitably approximate our 'pacing of change' condition (see de Block and Vis, 2018). The remaining difficulties were remedied by adjusting the initial indicators of 'endurance' and 'pace of change' until properly codable and yet theoretically valid operationalizations were arrived at (cf. Basurto and Speer, 2012).

Table 3. Operational coding rubric for the three design factors 'inclusivity', 'pace' and 'innovation'

Fuzzy set value	*Process Inclusivity*	*Pace of Policy Change*	*Degree of Innovation*
Fully out (0)	Few or no societal actors affected by the policy were *meaningfully informed or consulted* in the policy process	Change adopted in series of steps *over two terms* of government	All or most of the key elements of the reform/policy/program were *imitated*
More out than in (0.33)	Societal actors affected by the policy were *informed* throughout the policy process	Change adopted in series of steps *between one to two terms* of government	Most of the key elements of the reform/policy/program were *imitated*; few were *invented*
More in than out (0.67)	Societal actors affected by the policy were *consulted* throughout the policy process	Change adopted in series of steps/stages *within one term* of government	Most of the key elements of the reform/policy/program were *invented*; few were *imitated*

Fully in (1)	Societal actors affected by the policy *actively collaborated* with policymakers throughout the policy process	Change adopted and/or implemented in *one fell swoop*	All or most of the key elements of the reform/policy/program were *invented*

Source: The Author

To arrive at the outcome(s) of interest, we took the minimum value across different combination of success 'sets'. The minimum membership score, in effect, indicates the degree of membership of a case in a combination of sets (de Block and Vis, 2018). Its use follows 'weakest link reasoning' (Ragin, 2000). This means that for a case to be considered both programmatically *and* politically successful, it needs to have scored a minimum of 0.67 across both dimensions.[2] We were first interested in identifying pathways to programmatic and political success. This minimum value strategy was again used for the programmatic, political and process success outcome, as well as the programmatic, political, process and endurance success outcome. By doing so, we assert that success is not binary, or even a one-dimensional spectrum. Instead, it is multi-dimensional. By identifying and measuring multiple outcomes through the intersection of different success dimensions, we assert that a policy is considered successful when it scores *at least* 0.67 across the dimensions that are taken into consideration. The advantage of this method is that it acknowledges and accounts for the possibility that there is more than one formulation of success, and allows for further (future) theorizing on the different mechanisms that might underlie the manifestation of these different "routes" to success. The disadvantage of this approach, however, is that it may increase complexity. Recognizing the multidimensionality of success and allowing for different combinations of conditions to produce a "success" necessarily complicates any effort at causal or explanatory theory. Our goal here, though, is not to produce such theory.

Identifying routes to policy success: findings

The number of possible configurations generated in a QCA truth table is determined by the number of conditions used. The general rule is 2^k, where k is the number of conditions. As we have three conditions, the truth table (see table 4) shows eight logically possible configurations of those conditions. The technique then matches the cases included in our sample to the different combinations of conditions.

Table 4. Truth table for Policy Successes (n=31)

Conditions				
Process Inclusivity	**Pace of Change**	**Degree of Innovation**	**Consistency**	**Cases**
1	0	1	0.85	Bolsa Familia, UK Tobacco Control, NZ Early Childhood, NZ Whanau Ora, AU HIV/AIDS

2 — We applied QCA and SetMethods packages using R software (Dusa, 2019) to run the analysis.

1	1	0	0.83	Dutch Delta, AU Plain Packaging, AU GST
1	0	0	0.82	Finnish Education, Montreal Protocol, Marvellous Melbourne, NZ Waitangi, AU Competition Policy
1	1	1	0.80	GI Bill, AU Child Support
0	0	0	0.78	Danish Urban Planning, German Labour Market
0	1	1	0.63	NZ KiwiSaver, AU HECS, AU Gun Control, NZ Fiscal Responsibility
0	0	1	0.63	Singapore Healthcare, NHS Waiting Times, Estonia e-gov, Alameda Corridor, AU Water Markets, NZ Accident Comp
0	1	0	0.56	Norwegian Future Fund, AU Medicare, AU GFC Stimulus, NZ Economic Turnaround

Source: The Authors

In our analysis, we investigated three outcomes of interest, each being a different combination of the four success criteria. The first outcome of interest we considered was Programmatic and Political (PP) Success. The fsQCA analysis finds that there are essentially two routes which, if followed, are sufficient for leading to this type of success. The first route is to ensure the policy design process is inclusive. In other words, if policymakers want to achieve both programmatic *and* political success, then they must effectively consult and collaborate with affected societal actors throughout the policy process. The second route to PP Success results from a combination of a slow pace of change and a low degree of innovation. This pathway shows that policies which build on or are in line with previous efforts and are adopted slowly over a series of steps can achieve programmatic and political success.

It should be noted that either of these two pathways are considered sufficient, but not necessary, for producing PP Success. In fuzzy-set theory, logical AND (designated as *) refers to the combination of sets. Logical OR (marked as +) refers to the intersection of sets. The presence of a condition is denoted in upper-case, while the absence of a condition is in lower-case. In fuzzy-set notation, the so-called solution term of our analysis of PP success is represented as:

PROCESS INCLUSIVITY + pace of change * degree of innovation => PP (coverage: 0.73; consistency = 0.79)

However, as Tables 1 and 2 have shown, there is more to policy success than programmatic

and political achievement. To enrich the assessment of success incorporated in the analysis, we focused on two additional outcomes of interest: Programmatic, Political and Process (PPP) and a 'full Monty' Programmatic, Political, Process and Endurance (PPPE). When these were added to our analysis, it turned out that while the routes leading to both these new outcomes were identical, there were notable differences between these routes and those that produce the original PP outcome.

Again, there were two routes. The first is that of an inclusive process, fast pace and a low degree of innovation. The second again entails an inclusive process, but this time combined with a slow pace and high degree of innovation. In fuzzy-set notation, the solution term for these routes is represented as:

> PROCESS INCLUSIVITY*PACE OF CHANGE*degree of innovation +
> PROCESS INCLUSIVITY*pace of change*DEGREE OF INNOVATION => PPPE
> (coverage: 0.55; consistency: 0.86)

What does this teach us? Clearly *inclusive design* processes are pivotal if multi-dimensional (and thus enduring) success is to be achieved. Interestingly enough, this finding is based on a case sample that is skewed towards cases from majoritarian democracies, where inclusive consultation traditionally is not strongly embedded in the political culture of 'winner takes all'. Perhaps Lijphart (1999) did have a point in his 'evidence-based' extolling of the virtues of consensual democracies – where broader inclusion of groups, voices and interests is not a matter of choice but a political necessity.

Though pivotal, our findings indicate that inclusivity is not the only important factor. Variations in the degree of innovation sought by policymakers also determine which design route needs to be taken. To be considered successful on all four dimensions, policies that are low on innovation are best driven at a relatively fast pace (but through an inclusive process – which presumes policy designers need to adopt mechanisms for organizing broad consultation and perhaps co-design that do not lead to highly protracted deliberation and thus slow down the pace of the policy process). In contrast, if a highly innovative policy approach is to be successful, ample time must be taken for it to be developed inclusively.

To briefly illustrate both routes in practice, we turn to two examples from our case set: the 2008 Dutch Delta Reforms and Brazil's *Bolsa Família*.

Dutch Delta reforms: an inclusive, fast paced, and imitative policy design

Reform of the Dutch Delta approach illustrates the first 'path to success' identified by our analysis (based on Van Buuren, 2019). The review and revision of the existing water management policy that took place in the Netherlands during the latter 2000s was (1) inclusive, (2) fast-paced, and (3) highly imitative. Following increased awareness of the risks of climate change, the Second Delta Committee was created in 2008 to begin inclusive and collaborative deliberation on how to reform existing and internationally renowned Dutch Delta management programs. Those reforms passed in 2010 and revised the existing program to ensure the Dutch would remain safe from the encroaching sea in future decades.

Stakeholders from multiple levels of government, the private sector, and interest groups, as well as policymakers and bureaucrats, were included in the decision-making process or were invited to join advisory committees. The revised approach implemented in 2010 built upon and respected existing institutions and processes. Much about Dutch water policy was kept the same, and policy changes made were small and mostly voluntary. The second era of the Dutch

Delta approach self-consciously adopted a collaborative and piecemeal approach to adaptation based upon but iteratively complementing pre-existing policies. This way of designing the policy process strongly contributed to the legitimacy and authority of the ensuing policy reforms.

Bolsa Família: an inclusive, slow paced, and innovative policy design

The Brazilian conditional cash transfer (CCT) program *Bolsa Família* illustrates the second 'path to success' identified by our analysis (based on Paiva et al, 2019). Design and passage of this policy was (1) inclusive, (2) slow-paced, and (3) highly innovative. Currently reaching over 20% of the Brazilian population, CCTs have achieved significant results and *Bolsa Família* has been widely recognized as a major policy success.

Following a new democratic constitution in 1988 and increased decentralization, a debate took place among politicians and academics about the best policy to achieve poverty reduction. Reflecting the advice of academics, the first CCT model of social support was implemented in 1995 by two Brazilian municipalities. The highly innovative policy soon spread to other municipalities in the country. Federal implementation of CCT programs did not occur until 2001 under the Cardoso administration, which was three years after the Mexican government had sent a delegation to Brazil to study local-level CCTs and subsequently implemented their own national CCT program, *Progresa*. After years of experimentation at the local-level, and months of deliberation at the federal level, fragmented CCT programs were consolidated into one national program under the Lula da Silva administration in 2003: *Bolsa Família*. Enduring programmatic, political, and procedural success was achieved in this case by an inclusive and deliberative design that allowed highly innovative policy ideas to take root and weather critical scrutiny and skepticism over several years.

Conclusions: towards designing for success

In this paper we asked the grand question: whether policy success can be designed (and if so, how). That question obviously cannot be answered comprehensively on the basis of just one comparative study. We do hope, however, to have demonstrated some progress in moving towards a better grasp of the issues involved. Firstly, we offer Table 1 as an assessment tool, which analysts can use to gauge different dimensions and levels of policy success. Adding the 'endurance' criterion to McConnell's (2010) trichotomy of programmatic, process and political has, we believe, gone some way towards a more temporally sensitive understanding of success. A limitation of the assessment method used in this article has been its static nature (one summative assessment of policy histories spanning years and sometimes decades). Much can be gained by applying the assessment tool at different points in a policy's trajectory, creating a more 'film-like' grasp of how levels of success across the different dimensions develop through time. Doing so can generate important questions on how and why such variations can occur, and what design practices can contribute to success becoming achieved earlier, more consistently and across more dimensions.

Secondly, comparing and contrasting the four dimensions of success in the empirical analysis has highlighted the relevance of Bovens et al's (2001) observation that programmatic and political success do not necessarily align. Several cases in our sample exhibited 'mismatches': considerable programmatic achievements not matched by deep, wide and stable support (e.g. New Zealand's fledgling Whanau Ora policy of taking the Maori extended family unit as the locus for social policy interventions), or politically (temporally) popular and willed polices whose programmatic impact was relatively ephemeral (e.g. Tony Blair's 'cutting the wait at the NHS'

initiatives). Similar mismatches exist between programmatic (or political) and endurance assessments (policies 'peaking' at one point in time but being unable to be consolidated across changes of government, fiscal conditions and departures of key sponsors). Studying these mismatches and contrasting them with cases of more consistent and comprehensive success using structured and focused comparison of key decisions and processes should yield insights about what design practices may be conducive to less conflicted and more enduring forms of success.

Thirdly, the present study is not without methodological limitations. One stems from our reliance on secondary analysis of narrative case studies written by a multitude of authors. All authors of the original cases worked with (a version of) Table 1 as a central tool for structuring their narratives, but they were not instructed to explicitly cover the three design factors presented in this paper. In other words, coverage of all conditions of interest has been somewhat uneven, complicating the coding process. Clearly, a 'triple hermeneutic' has been at work in the present research's design: we were interpreting the authors' interpretations of the case actors' interpretations of these policies. Future comparative studies of policy success should find ways of reducing the potential 'noise' that can occur in the present set-up.

Another limitation is a consequence of our decision to select a population of 'success cases.' Selecting a sample of cases based on the outcome (success) may raise eyebrows among those committed to constructing samples based on (maximal) variation on that dependent variable. As discussed above, however, we have been conscientious in our inferences, and are therefore on firm ground by 'selecting on the dependent variable.' By constructing four dimensions and levels of success and coding the cases accordingly, we generated sufficient variation to allow the fsQCA-analysis to yield valid inferences and meaningful results. We have uncovered the (combinations of) conditions conducive to policy success—a result of real value to the literature. To generate conclusions about the average or marginal contribution of different factors to outcomes, a different research design should be employed. To do this, future research could endeavor to contrast highly (comprehensively and enduringly) successful cases with 'others' (comprehensive and chronic policy failures), holding constant policy areas and jurisdictions (Bovens et al, 2001 present a good example).

In sum, this study provides analysts with concepts and tools for achieving a more 'granular' understanding of the layered meanings of the term 'works' in the phrase uttered so often by political decision makers and policy designers that "we should adopt 'what works'". It also advances our knowledge of what principles and practices of policy design – specifically around inclusion, pacing, and innovation – can help or hinder the achievement of policy success, casting light on the nature of the 'what' in that same phrase. This is only a modest beginning, which we hope will inspire others to build upon and test the impact of other configurations of design conditions so that a decade from now the field will possess an evidence-based catalogue of 'tried and tested' design strategies.

Bibliography

Alford, J. (2009). Engaging Public Sector Clients: From Service-Delivery to Co-Production, London: Palgrave.

Basurto, X., & Speer, J. (2012). Structuring the calibration of qualitative data as sets for qualitative comparative analysis. *Field Methods, 24*(2), 155-174.

Bennett, C. J. (1991). What is policy convergence and what causes it?. *British Journal of Political Science, 21*(2), 215-233.

Bobrow, D., & Dryzek, J. (1987). *Policy Analysis by Design*, Pittsburgh, PA: Pittsburgh University Press.

Boin, A., McConnell, A., & 't Hart, P. (2009). Towards a theory of crisis exploitation: Political and policy impacts of framing contests. *Journal of European Public Policy, 16*(1), 81-106.

Bovens, M., 't Hart, P., & Peters, B.G. (Eds.)(2001). *Success and Failure in Public Governance: A Comparative Analysis*. Cheltenham: Edward Elgar.

Bovens, M., 't Hart, P., & Kuipers, S.L. (2006). The politics of policy evaluation. In M. Moran, M. Rein, B. Goodin (Eds.), *Oxford Handbook of Public Policy* (pp. 319-335). Oxford: Oxford University Press.

Brady, H. E., & Collier, D. (2010). *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Plymouth: Rowman & Littlefield.

Bryson, J. M., Crosby, B.C., & Bloomberg, L. (2014). Public value governance: Moving beyond traditional public administration and the new public management. *Public Administration Review, 74*(4), 445-456.

Busetti, S., & Dente, B. (2018). Designing multi-actor implementation: A mechanism-based approach. *Public Policy and Administration, 33*(1), 46-65.

Capano, G., Howlett, M., Ramesh, M., & Virani, A. (Eds.)(2019). *Making Policies Work: First- and Second-Order Mechanisms in Policy Design*. Cheltenham: Edward Elgar.

Compton, M.E., & 't Hart, P. (Eds.) (2019). *Great Policy Successes*. Oxford: Oxford University Press.

Cyert, R.M., & March, J.G. (1963). *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice Hall.

De Block, D., & Vis, B. (2018). Addressing the challenges related to transforming qualitative into quantitative data in qualitative comparative analysis. *Journal of Mixed Methods Research*, 1558689818770061.

De Francesco, F., & Maggetti, M. (2018). Assessing disproportionality: Indexes of policy responses to the 2007-2008 banking crisis. *Policy Sciences, 51*(1), 17-38.

Dror, Y. (1986). *Policymaking Under Adversity*. New Brunswick: Transaction.

Dror, Y. (2001). *The Capacity to Govern*. London: Frank Cass.

Dusa, A. (2019). *QCA with R: A Comprehensive Resource*. Springer International Publishing.

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

Geddes, B. (2003). *Paradigms and Sandcastles: Theory Building and Research Design in Comparative Politics*. Ann Arbor, MI: University of Michigan Press.

Gouillart, F., & Hallett, T. (2015). Co-creation in government. *Stanford Social Innovation Review*. Retrieved from https://ssir.org/articles/entry/co_creation_in_government.

Grace, F.C., Meurk, C.S., Head, B.W., Hall, W.D., Harris, M.G., & Whiteford, H.A. (2017). An analysis of policy success and failure in formal evaluations of Australia's national mental health strategy (1992–2012). *BMC Health Services Research, 17*(374), 1-15.

Hacker, J. S. (2004). Privatizing Risk without Privatizing the Welfare State: The Hidden Politics of Social Policy Retrenchment in the United States. *The American Political Science Review, 98*(2), 243-260.

Hayes, R. (2017). Incrementalism and public policy. In: *Oxford Encyclopedia of Politics*. https://doi.org/10.1093/acrefore/9780190228637.013.133

Heckman, J. A. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement, 5*(4), 475–492.

Heifetz, R., Linsky, M., & Grashow, A. (2009). *The Practice of Adaptive Leadership*. Cambridge: Harvard Business Review Press.

Hogan, J., & Feeney, S. (2012). Crisis and policy change: The role of the political entrepreneur. *Risks, Hazards and Public Policy, 3*(2), 1-24.

Hood, C., & Margetts, H. (2007). *The Tools of Government in the Digital Age*. Basingstoke: Palgrave Macmillan.

Hough, M., Jackson, J., Bradford, B., Myhill, A., & Quinton, P. (2010). Procedural Justice, Trust, and Institutional Legitimacy. *Policing, 4*(3), 203-210.

Howlett, M. (2012). The lessons of failure: learning and blame avoidance in public policy-making. *International Political Science Review, 33*(4), 539–555.

Howlett, M. (2014). From the 'old' to the 'new' policy design: design thinking beyond markets and collaborative governance. *Policy Sciences, 47*(2), 187-207.

Howlett, M. (2019). *Designing Public Policies*. London: Taylor and Francis.

Howlett, M., & Mukherjee, I. (Eds.)(2018). *Routledge Handbook of Public Policy Design*. London: Routledge.

Kay, A., & Boxall, A.M.. (2015). Success and failure in public policy: Twin imposters or avenues for reform? Selected evidence from 40 years of health-care reform in Australia. *Australian Journal of Public Administration, 74*(1), 33-41.

King, G., Keohane, R.O., & Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

Klijn, E.H., & Koppenjan, J. (2016). *Governance Networks in the Public Sector*. London: Routledge.

Lijphart, A. (1999). *Patterns of Democracy*. New Haven: Yale University Press.

Lindblom, C.E. (1979). Still muddling. Not yet through. *Public Administration Review, 39*(6), 515-526.

Linder, S.H., & Peters, B.G. (1987). A design perspective on policy implementation: The fallacies of misplaced prescription. *Policy Studies Review, 6*(3), 459-475.

Linder, S.H., & Peters, B.G. (1988). The design of analysis or the analysis of design. *Review of Policy Research, 7*(4), 738-750.

Luetjens, J., Mintrom, M., & 't Hart, P. (Eds.)(2019). *Successful Public Policy: Lessons from Australia and New Zealand*. Canberra: ANU Press.

Mahoney, J., & Goertz, G. (2006). A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research. *Political Analysis, 14*(3), 227–249.

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

March, J. (1991). Exploration and exploitation in organizational learning. *Management Science, 2*(1), 71-87.

McConnell, A. (2010). *Understanding Policy Success*. London: Palgrave.

Nielsen, S.B., Turksema, R., & van der Knaap, P. (2015). *Success in Evaluation: Focusing on the Positives*. New Brunswick: Transaction.

Newman, J. (2014). Measuring policy success: Case Studies from Australia and Canada. *Australian Journal of Public Administration, 73*(2), 192-205.

Olavarria-Gambi, M. (2018). Policy failure revisited: Conceptual implications from the Chilean case of Transantiago. *Administration and Society*. https://doi.org/10.1177/0095399718786878

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

Paiva, L.H., Cotta, T.C., & Barrientos, A. (2019). Brazil's Bolsa Família programme. In M. Compton & P. 't Hart (Eds.), *Great Policy Successes*. Oxford: Oxford University Press.

Peters, B.G. (2018). *Policy Problems and Policy Design*. London: Edward Elgar.

Peters, B.G., Capano, G., Howlett, M., Mukherjee, I., Chou, M-H., & Ravenet, P. (2018). *Designing for Policy Effectiveness*. Cambridge: Cambridge University Press.

Posen, H.E., Keil T., Kim, S. & Meissner, F. (2018). Renewing research on problemistic search: A review and research agenda. *The Academy of Management Annals, 12*(1), 208-251.

Ragin, C.C. (2000). *Fuzzy-set Social Science*. Chicago, IL: University of Chicago Press.

Ragin, C.C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: The University of Chicago Press.

Rihoux, B., & Ragin, C.C. (Eds.)(2009). *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Los Angeles, CA: SAGE.

Rogers, E.M. (2003). *Diffusion of Innovations*. New York, NY: The Free Press.

Rothmayr Allison, C., & Saint-Martin, D. (2011). Half a century of "muddling": Are we there yet? *Policy and Society, 30*(1), 1-8.

Schneider, C.Q., & Wagemann, C. (2012). *Set-Theoretic Methods for the Social Sciences*. Cambridge: Cambridge University Press.

Torfing, J. (2019). Collaborative innovation in the public sector: the argument. *Public Management Review, 21*(1), 1-11.

Torfing, J., & Triantafillou, P. (2016). *Enhancing Public Innovation by Transforming Public Governance*. Cambridge: Cambridge University Press.

Van Assche, K., Beunen, R., & Duineveld, M. (2012). Performing success and failure in governance: Dutch planning experiences. *Public Administration, 90*(3), 567-581.

Van Berkel, F., Ferguson, J., & Groenewegen, P. (2016). Once bitten, twice shy: How anxiety and political blame avoidance cause a downward spiral of trust and control in the aftermath of failed public projects. *Administration and Society, 51*(4), 545-580.

Van Buuren, A. (2019). The Dutch Delta: The successful reinvention of a policy success. In M. Compton & P. 't Hart (Eds.), *Great Policy Successes*. Oxford: Oxford University Press.

Van den Bos, K., van der Velden, L., & Lind, E.A. (2014). On the role of perceived procedural justice in citizens' reactions to government decisions and the handling of conflicts. *Utrecht Law Review, 10*(4), 1-26.

Vis, B. (2012). The Comparative Advantage of fsQCA and Regression Analysis for Moderately Large-N Analysis. *Sociological Methods and Research, 41*(1), 168-198.

Weible, C., & Cairney, P. (Eds.) (2018). Practical lessons from policy theories. *Special issue of Policy Politics, 46*(2), 183-343

Appendix

The core features and brief case summaries of the 33 cases under investigation are presented in Appendix 1. A description of the coding protocol and justifications are reported in Appendix 2 (further coding justifications are available upon request). All the material used in the coding of these cases can be found in Compton and 't Hart (2019) and Luetjens et al. (2019).

Appendix 1: 33 Policy Success Cases (reported and sourced in Compton and 't Hart, 2019 and Luetjens et al, 2019)

Overview of cases

Feature	Cases
Majoritarian democracies (two-party systems)	24
Consensus democracies (multi-party systems)	6
Federal systems	16
Unitary systems	16
Global governance	1
National governance	29
Local/urban governance	3

Note: Brazil's *Bolsa Familia*, Singapore's healthcare performance and the Montreal Protocol are not included in the majoritarian/consensus classification. The Montreal Protocol is not included in the federal/unitary classification.

Brief case summaries

Brazil's Bolsa Familia scheme – The world's largest conditional cash transfer scheme that lifted millions out of extreme poverty.

Healthcare performance in Singapore – How Singapore has drastically increased its health outcomes while keeping costs low.

Cutting waiting times in the NHS – Combination of resolute political leadership and judicious policy

analysis that enabled Britain's National Health Service to process its millions of clients much quicker.

UK's tobacco control regime – How the UK designed and implemented innovative policies which framed tobacco as a health concern to successfully build support around the initially unpopular tobacco ban.

The GI Bill – How the United States provided social support to soldiers returning from the Second World War to ensure macro-economic security, and had the unintended consequence of building social capital.

Finland's secondary school system – How a small nation on Europe's northern periphery's school system became a global brand in 'how to do public education'.

Estonia's digital transformation – How a post-communist state forged a global reputation as a leader in digital government.

The Los Angeles region's Alameda rail corridor project – A balanced governance and creative financing arrangement transforming a tangled web of rail lines into a single corridor that relieved traffic congestion and reduced air and water pollution in the Los Angeles region.

The new Dutch Delta strategy – How a nation in which two-thirds of the population live below the current sea level secures its future by reinventing its famed water management strategy so as to enable proactive and creative adaptation to the effects of climate change.

Copenhagen's Five Finger Plan – How the Danish capital successfully avoided urban sprawl and overly sense and chaotic urbanization through early adoption and sustained adaptation of a comprehensive urban planning regime.

Norway's Petroleum Fund – How Norway's policymakers purposefully dodged the bullet of the 'resource course' and channeled its oil revenues into what has become the world's biggest national pension fund.

Germany's labor market reforms – How Europe's biggest but notoriously rigid and sluggish post-reunification economy was lifted into the economic powerhouse it has since become.

The Montreal Protocol – How the world managed to negotiate and implement a global regulatory regime that helped the stratospheric ozone layer recover from the damage sustained by decades' worth of ozone depleting substances.

'Marvellous Melbourne' – How the once staid state and struggling capital of Victoria, Australia transformed itself into a cosmopolitan metropolis named 'The World's Most Liveable City' six times in a row (from 2011 to 2017) by The Economist's Intelligence Unit.

Australia's response to HIV/AIDS – How an effective national policy response to HIV/AIDS evolved across challenging social, political, epidemiological, medical and generational contexts.

Australia's Higher Education Contribution Scheme – The first national income-contingent university tuition fee loan program that saw enrolments triple and inspired similar loan schemes around the world.

Australia's economic crisis management – Massive pre-emptive macro-economic policy response to the Global Financial Crisis in 2008 which led to the avoidance of recession but caused heated debate about its methods.

Australia's Child Support Scheme – Increasing the proportion of children of separated parents receiving financial support as well as increasing the contributions paid to government by separated parents.

Australian water markets – A transformative and widely supported change to water allocation processes in the Murray-Darling Basin

Australia's National Competition Policy – A sustained and impactful program of product market liberalization hailed as an example of successful collaboration between Australia's federal and state governments.

Australia's gun control reform – Following a mass shooting in Tasmania, the federal government swiftly implemented a national firearm policy that turned Australia into a world leader in the prevention of armed violence.

Australia's Goods and Services Tax – A once highly contentious tax policy reform that was well-designed, effectively implemented and enjoyed broad political and public support.

Australia's Medicare – The foundation of Australia's universal health care system that evolved into a widely popular institution.

Australia's avoidance of financial institutions' collapses in 2008-9 – Good luck, a sound regulatory environment and the adherence of banks to 'boring but safe' business models allowed Australian financial institutions to dodge the bullet of the Global Financial Crisis.

Australia's Tobacco Plain Packaging – The first nation to implement tobacco plain packaging laws, effectively encouraging current smokers to quit and dissuading potential new smokers from starting.

New Zealand's economic turnaround – A country at the brink of economic collapse in the 1980s transformed its fortunes through a radical, consistent and impactful suite of reform strategies.

New Zealand's Accident Compensation Scheme – Unique universal accident-insurance system, based on compulsory contributions into a state monopoly with no right to sue.

Nuclear Free New Zealand – How despite major power pressures to desist, a small state adopted and sustained a nuclear free stance that has attracted widespread support across the political spectrum.

Treaty of Waitangi Settlements – New Zealand's policy to address, negotiate, and settle historical grievances arising from the Crown's Treaty of Waitangi breaches.

New Zealand's 1994 Fiscal Responsibility Act – Engraining the principle of fiscal responsibility in New Zealand budgetary practices and constitutional arrangements.

New Zealand's Early Childhood Education Pathways – Comprehensive policy plan that greatly increased young children's participation in care and education.

New Zealand's KiwiSaver pension scheme – Delivering one of the world's lowest rates of poverty for the population aged 65 and over, while offering citizens flexibility to react to financial events throughout their lives.

New Zealand's Whānau Ora – Implementing values-based innovative solutions for Māori communities, building their capacity for self-management.

Appendix 2: Coding Protocol and Exemplars

As discussed in the text, each author was responsible for coding a third of the cases. As these cases emerged from two broader collaborative projects, each author had extensive familiarity with the contents of each case, having been closely involved in the production phase. Each author had been involved in workshops where the case authors presented drafts of their cases. The authors then discussed and edited these drafts for inclusion in the two broader collaborative (edited volume) projects. The case authors were identified and selected as a result of their extensive experience either directly with the design and implementation of the policy under consideration, or given their sub-

ject matter expertise. The case authors drew on previous interview material, their published works, and/or their direct experience in developing the case chapters. The case authors were given guiding questions for their case analysis (these can be found in Compton and 't Hart (2019) and Luetjens et al. (2019)). The result of these guiding questions is that the diverse set of cases are explored with a common set of reference points. This approach offers many opportunities for comparisons to be drawn out from across the whole set.

This common set of reference points facilitated the coding of these cases for the purposes of this article. Below, we offer three examples of our coding decisions. To justify the score, the authors relied heavily on supporting material found in each case. As noted in the text, the authors first scored a third of the cases and then scored an additional set of cases that had previously been scored by a different author. The scores were then compared to identify discrepancies in the coding decisions. Few discrepancies were found. The main issue that emerged was during the initial round of coding. This had to do with the way in which the conditions and outcome were initially conceptualized. To rectify this, and to ensure that all cases (except Australia's response to the Global Financial Crisis) could to be coded, the authors revisited the operationalization of their conditions and outcome(s). The operationalization(s) arrived at in Tables 1 and 2 represent the outcome of this effort

Finnish Comprehensive Education

Dimension	*Score (1-4)*	*Explanation/consideration*
Programmatic	4	Performance of the Finnish comprehensive education system is demonstrated in at least two dimensions: "economic growth aided by human resources; and…upward social mobility education affords." Evidence points to success in both economic and equality terms. Since the 1990s, the Finnish system "has basked in international glory, being called one of the best in the world. This reputation is largely due to the successful performance of Finnish teenagers in the Programme for International Student Achievement (PISA), run by the Organisation for Economic Co-operation and Development (OECD)."
Process	4	The process of adopting the education system was long and deliberative. Drafting the legislation was done by "independent, broadly politically representative, and expert-based committees served as *ad hoc* organs in producing reports and drafting laws." The eventual compromise on design of the program was based on expert research and deliberation.
Political	4	Initial implementation of the program resulted in a lasting compromise around which different parties and interests eventually agreed. Stakeholders and policymakers associated with the system receive external (international) recognition, as well as domestic political rewards.

Endurance	4	"Despite repeated criticism, institutional frame survived and is now a recognized and almost unchallenged part of the Finnish education landscape." Continuation of the program was in part due to the frequency of coalition governments in Finland, which almost always included some parties from the previous administration in the new government. What changes have occurred were minor calibrations of existing instruments. "…[I]n the last sixty years there has been no initiative to change the fundamental principles of the original policy establishing the comprehensive school system (Kauko et al. 2015; Simola et al. 2017)."
Process Inclusivity	4	Over the decades of continued implementation, ideologically heterogeneous coalitions continued the program. The process of designing the program in the first place was deliberative and inclusive of multiple differing viewpoints. The process "could be seen as fair in the sense that the opinions of the opposition were considered and deliberated in parliamentary decision making. However, disputes arose during the implementation phase concerning what had actually been agreed to in the policy."
Pace of Change	1	"The political process leading to the complete reorganization of the formerly bipartite school system and the establishment of the comprehensive school and its implementation took more than three decades—and even longer if we track the origin of some of its constitutive ideas. " "Incremental advances eventually resulted in a critical juncture in which the comprehensive school was created in the late 1960s."
Degree of Innovation	3	The author does not explicitly discuss the innovativeness of the 1967 government bill that would create the comprehensive school committee. The degree of attention from external/international actors suggests that the Finnish system was innovative and remains unique.

Australian Higher Education Contribution Scheme

Dimension	Score (1-4)	Explanation/consideration
Programmatic	3	HECS achieved its intended social outcomes by successfully facilitating expansion of the higher education and graduate population without compromising access. It's role in facilitating an affordable and equitable means of funding higher education has created considerable public value: "The tide of expanded participation lifted all boats." However, the costs on students are still quite considerable. The debate as to the correct level of student charges versus public funding is ongoing. Determining how costs should be distributed has proved problematic.

Process	2	Highly centralized process; use of handpicked committees and external experts; the policy design process "occurred quickly, providing limited time or chance for opponents to disagree." At that time, the innovative scheme had no international precedent: "it was untested and there was no empirical evidence that it would succeed." There was no published research paper setting out the idea that the government could refer to.
Political	2	Both major parties now view HECS as an essential component of the higher education system. The model has been exported around the world. The architect, Bruce Chapman, has received international and national recognition and accolades. However, the Coalition (the Opposition party) was initially opposed, only coming around to the idea 5 years later. Similarly, the ALP Party Platform was initially opposed to fees – national student and staff unions lobbied the ALP Caucus. In its early years, it took time and resources to defend.
Endurance	3	Following the 1996 election, the Howard government "chose its retention." The goals of the policy have largely remained intact. However, HECS has been subject to changes often driven by the political attraction of budgetary savings. While HECS has been resilient and endured for the past 30 years, the rising stock and cost of debt and calls to reform the higher education sector pose challenges.
Process Inclusivity	2	"The Government was faced with the task of consulting with and convincing stakeholders and the public of the merits of the proposal." The ACTU was brought on board after the development of the proposal. The policy decision process was set and controlled by the education minister: He handpicked committee members and secretariat staff. "The consultation process was brief, but it was particularly well focused."
Pace of Change	4	The process moved quickly with HECS becoming law only 18 months after the proposal was developed, well within one term of government.
Degree of Innovation	4	Bruce Chapman's report included the novel idea of an income contingent charge to be repaid through the tax system following graduation, something no other country had previously implemented. "HECS was introduced on a Greenfield policy site, policy constraints and conflicts did not exist to divert the development process."

German Labour Market

Dimension	Score (1-4)	Explanation/consideration
Programmatic	3	The reforms were successful in that they "enabled the effective tackling of unemployment even during the worst recession in post-war history which acted as a major test for the economy's robustness (Rinne and Zimmermann 2013)." Macroeconomic indicators suggest the policies were successful in meeting goals, but evidence of the effectiveness of compensatory measures is less clear.
Process	4	Choice of policy mixes were also well suited to the existing economic model in Germany and have been updated (through continuous trial and error) in the face of changing circumstances. The designers—a 'strategic centre'—relied heavily on the experiences of similar countries in making evidence-based policy decisions. Also, procedural success is evidenced in the Hartz Committee's ability to enhance problem-solving capacity and to circumvent a 'stalemate in corporatist policy making.'
Political	2	The passage and impacts of the Hartz reforms were highly contentious. "While business and employer associations as well as conservative and liberal parties supported the Hartz reforms, unions, social welfare organisations, leftist parties, and parts of the public criticised social cutbacks." However, "over time, both the inclusiveness (a process component) and breadth and depth of the societal legitimacy (a political component) of German labour market policies have improved." Today, "there is no societal consensus on the core objectives of labour market policies, and there is a widespread sense of injustice."
Endurance	3	The initial Hartz reform instruments are no longer in place, and have been replaced by subsequent reforms. The Hartz IV law has not been terminated or changed. These changes reflect evolution in economic circumstances in Germany, rather than changing principles or goals.
Process Inclusivity	1	As Spohr writes, "the composition of the Hartz Committee was not about inclusion and consensus but about expertise and a will to compromise." Several important societal actors were left out of the design stage. This process "reduced procedural justice since social partnership negotiations serve the legitimisation of government actions; governments especially incorporate trade unions into policy making and implementation for their own political support (Hassel 2009)."
Pace of Change	1	The first reforms were passed in 2002, with laws coming into effect between 2003 and 2005. The second wave of reforms were passed amid the great recession in 2009. Together these changes took place over more than 2 government terms.

Degree of Innovation	1	The Hartz Reforms relied heavily on the experiences of others. As Spohr writes, "policy mixes like the Hartz reforms – activating the unemployed, improving their employability and making low-skilled labour productive – had already been implemented quite successfully in social democratic Scandinavian welfare states such as Denmark and Sweden as well as in liberal Anglo-Saxon systems such as the UK and the US since the 1980s"