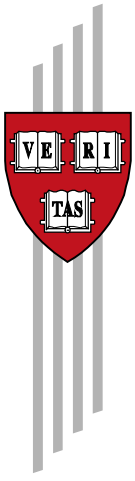
Successful Failure in Public Policy Work

Successful Failure in Public Policy Work

Matt Andrews

CID Faculty Working Paper No. 402
December 2021



Working Papers

Center for International Development at Harvard University

Successful Failure in Public Policy Work

Matt Andrews

Harvard Kennedy School

Abstract

It matters if public policies succeed in solving societal problems, but a dominant narrative holds that policies fail 'often'. A large-sample study discussed in this paper suggests that this is not accurate, however. The most common policy result in this study is more ambiguous—what I call 'successful failure'. Such result is achieved when a policy delivers enough low-level, short-term product to promise success, but ultimately (and repeatedly) fails to contribute to sustained high-level, long-term impact (addressing the problems citizens care about). Such 'successful failure' is endemic to public policy work, and a more pernicious result than outright failure: It allows policy design and implementation actors to associate with incomplete near-run success but insulate themselves from future failure (which they blame on factors and actors beyond their control) and simultaneously enjoy repeated demand for work (because problems are never really solved).

Public Policy Stories From the Field

The French health system has struggled with coordination problems and cost concerns for decades. These problems prompted an effort to adopt a national electronic health records policy in the early 2000s (Burnel 2018, Manard et al. 2019, and Séroussi and Bouaud 2016, 2018 and 2020). This effort delivered outputs like a new law in 2004 and a new IT system in 2011. Outside observers lauded these deliverables as examples for others to emulate (Grady 2012, IHE 2018). Unfortunately, however, they have had less impact than initially hoped; with few doctors or patients using the system.[1] As a result, coordination and cost problems persist (Graeve 2020).

A similar story comes from Brazil's education sector, where government has pursued many policies to improve quality over the last thirty years. These policies have delivered schools, training facilities and management protocols. Such outputs are considered 'best practice' in the education realm, credited with improving outcomes like school enrollment (Haddad et al. 2017, Horn 2002, Schwartzman 2005). Unfortunately, however, the policies have failed to improve student performance, with a recent study finding that, "[I]t would take Brazil more than 260 years to reach the OECD average proficiency in reading and 75 years in mathematics" (Raiser 2018).

Elsewhere, South Africa has long struggled with unemployment. In response, governments introduced five economic policies since 1994 (Aron et al. 2009, Hausmann and Andrews 2009, Thwala 2011, Naidoo and Maré 2015). Each policy delivered outputs—like welfare, skills development programs, laws, and infrastructure. Some outputs have been showcased as models for other governments to copy. Unfortunately, however, the outputs have not solved or contained the problem, and the unemployment rate grew from 20 percent in 1994 to 35 percent in 2021.[2]

Looking further south, Australia has struggled with the effects of climate change for decades. Governments have tried to address this problem through policies designed to reduce carbon emissions (Crowley 2017, 2021, Hudson 2019, Müller and Slominski 2017). These policies have yielded a plethora of schemes, commissions, plans, and legislation—including a landmark 2012

[1] Even with a new initiative in 2016, user coverage in 2018 was less than 20% of goals set for 2012 (Burnel 2018).
[2] See data at http://www.statssa.gov.za/?s=unemployment+rate&sitem=content and http://www.statssa.gov.za/?page_id=737&id=1&paged=2

law and carbon pricing mechanism that many observers considered a global best practice. Unfortunately, most of these initiatives (including the 2012 law) were short-lived or ineffective, and commentators now note that carbon emissions have not been decreasing as needed. As a result, the country's climate change problems seem to be worsening.

Introduction

I define public policies as the sets of activities governments undertake to address societal problems (problems that are recognized to affect collectives or to require a collective response).[3] The scale and scope of these activities has expanded significantly in recent generations, as more problems have been deemed societal by social, economic and political decision makers.[4] New problems on public policy agendas differ from country to country but include social issues like education, health and welfare, environmental challenges like climate change, and concerns related to globalization like migration and inequality. Such new problems are represented in my four stories from the field: Health care costs became a major issue in France only in recent decades; education equity and quality emerged as a major agenda item in Brazil from the 1990s; unemployment received increased attention from South African policymakers after the 1994 elections; the need for emissions control was only embraced from the 1990s onwards in Australia.

Millions of people are affected by the problems such policies are created to address, as is the legitimacy of governments crafting and implementing the policies. It is thus important to ask how regularly the policies succeed in containing or solving their focal problems. Unfortunately, a prominent narrative suggests that public policies 'often' fail in so doing, and some even claim that this "failure … is ubiquitous" (Fotaki 2010, Schuck 2014, Mueller 2020). If true, this claim deserves urgent attention from social decision-makers across the world—who need to look for better ways of finding solutions to societal problems. I do not see evidence of this attention, however, as public

[3] Non-governmental and private entities can also undertake such activities but are seldom the main actors, at least over the long term (where governments have both the responsibility and resources to do this work).

[4] The expansion of policy agendas reflects the expansion of democratic systems across the world, and the broadening of participation and representation in many of these systems. More problems seem to be deemed 'societal' when more—and more diverse—people and groups are given voice in their political processes. Relevant articles making this point in the context of the United States include Hill et al. (1995), Lott and Kenny (1999) and Besley and Case (2003).

officials across the globe continue to grow their policy agendas—actively and even uncritically—and hold that failure is not only uncommon but 'unacceptable' in their work.

There are reasons why I believe the attention to failed policies is not forthcoming. Most fundamentally, the 'failure happens often' claim is backed by limited evidence, with advocates typically referring to one case or a small sample of cases where policies resulted in unambiguous or even catastrophic failure. Few—if any—large sample studies exist that examine results in a broad set of policy work undertaken across governments, asking if (i) clear policy failures outnumber uncontested policy successes, and (ii) if those policies that yield more ambiguous results (which we cannot easily call successes or failures). Without such evidence of widespread failure, public officials will argue that they have no reason to question the expansion of policy work and face up to failure in their work (instead of dismissing it as unacceptable). And maybe they will be right.

This paper responds to the dearth of evidence on policy success and failure with a modest version of the kind of large sample study that is currently lacking. In it, I describe an analysis of the results of 999 policy engagements undertaken with World Bank financing in over 100 countries. Drawing on the World Bank's most reported 'outcome' evaluations ratings, I find that the bulk of this policy work is not actually deemed a failure. In fact, up to 74 percent of the policy engagements in the sample were considered successful in delivering short and medium-term outputs and outcomes. The World Bank actively quotes this statistic in its annual reports, indicating a high level of success in the policy work it supports (Behrens at al. 2018). This is not the full story, however. In less commonly reported ratings, World Bank evaluators also reveal that up to 91 percent of the policy engagements are not confidently expected to achieve sustained long-run impact (given fears that they will not reach or sustain results, captured in the 'risk to development outcome' indicator).

When combined, this data shows that most policy engagements often achieve low-level to mid-level programmatic success—mobilizing new activities, delivering outputs, and achieving some outcomes—but most policy engagements do not yield high-level success—sustained impact that contains or solves the societal problems that prompted policy work in the first place. I describe this common public policy result as 'successful failure': Where a policy engagement delivers

enough low-level, short-term product to promise success but ultimately (and repeatedly) fails to contribute to sustained high-level, long-term impact (addressing its focal problems).

In a concluding discussion, I note that all four of my stories from the field are examples of such 'successful failure': The French government has successfully produced a two-decade stream of laws, IT systems and more, but health care costs continue to rise; Brazil's government has delivered schools, syllabi, teacher training programs and more, but education quality is still low; South Africa's government has generated multiple plans to combat unemployment, but unemployment keeps growing; Australia's government crafted laws, reports and agencies to address emissions, but emissions remain high. Referencing these and other examples, I argue that 'successful failure' is endemic, and a much more pernicious result than outright (or even catastrophic) failure: It allows public officials to associate with low-level to mid-level, short-run success but insulate themselves from higher-level, long-run failure (which they blame on factors beyond their control) and enjoy repeated demand for work (because problems are never really solved). I finish the paper by highlighting the importance of identifying 'successful failure' as the dominant—and most problematic—result we see in public policy (rather than outright failure).

Public Policy: A Growing Agenda of Failed Activities?

As introduced, I view public policies as the activities governments undertake to address societal problems—social, economic, and other challenges that affect collectives or require collective response. As also mentioned, countries across the world are heavily dependent on such public policy work. This work accounted for 39 percent of Gross Domestic Product (GDP) in the average country in 2020, having expanded from 24 percent in 1970.[5]

The growing role of public policy reflects over a century of policy accumulation across the world (Adam et al. 2019). Some of this accumulation occurs when political leaders add new policies to address newly (or more seriously) recognized problems. This is arguably what occurred when the United States federal government expanded its work on social policy concerns like health, housing

[5] This is the share of government spending in GDP, with statistic based on the author's analysis of data from Our World in Data (Ortiz-Ospina and Roser 2020) and the IMF https://www.imf.org/external/datamapper/

and education between 1950 and 1980 (when the USA's share of government expenditure to GDP grew from 13 to 37 percent).[6] This kind of additive accumulation is happening in many countries now with respect to emerging challenges like climate change and global migration.

Policies also accumulate when governments layer new activities on top of established policy programs, increasing the policy footprint of such programs. Such layering is often motivated by the identification of new problems in existing policy domains. This has arguably occurred in the United States' federal government's education policy over recent generations where new activities have constantly been added to address shifting problems—like enhancing access for disabled communities in the 1970s, improving school quality in the 1990s, and ensuring equal education standards in the 2000s.[7] This kind of layering accumulation occurs all the time, reflected in the annual increases one sees in most sector and ministry budget allocations around the world.

Given these forces of policy accumulation and the subsequent growth in the role of policy work in modern society, it makes sense to ask a blunt question: How regularly do public policy initiatives succeed or fail—especially when those initiatives address new challenges?

This is an important question, as millions of people are affected by the emergent problems public policies are created to address, and these policy responses draw on limited social resources (time, talent, money and more) that we cannot afford to waste. Additionally, observers argue that the regularity of policy success and failure affects the legitimacy of democratic polities (with repeated failures undermining democratic systems and repeated successes bolstering such).[8] Given this, it matters if (and how regularly) public policies succeed and fail, and we should know the answer. It is not easy to analyze patterns of public policy success and failure, however, partly because there are many different—and conflictual—views on what public policy success (and failure) even means

[6] Growth in these policy arenas led to the creation of five new Cabinet positions: Health and Human Services (1953), Housing and Urban Development (1965), Transportation (1967), Energy (1977), and Education (1979).

[7] These focal points are reflected in, among other policy activities, the 1954 National Defense Education Act, 1986 Handicapped Children's Protection Act, 1994 Improving America's Schools Act, and 2002 No Child left Behind Act. All were major policy shifts to address changing understandings of problems facing the education community.

[8] In his analysis of Swedish municipal legitimacy, for instance, Anders Hanberger (2003, 270-271) finds that "legitimacy increases through 'problem-oriented' and 'problem-effective' policies" (where governments tackled problems their citizens cared about and managed to address such with some level of effectiveness). He also notes, however, that other factors influence legitimacy, noting particularly that, "under pre-democratic regimes, as well as during periods of centralized democracy, the support for, and confidence in, legality seems to be the dominant source of legitimacy."

(Compton and t'Hart 2019, Marsh and McConnell 2010, McConnell 2010).[9] What some might call a success could be seen as failure by others, and their views may be irreconcilable.[10]

Even with such fundamental ambiguity, however, I have observed a dominant narrative emerging in answer to the question posed. It is simple, and negative, and holds that most public policies fail. This narrative is reflected in a 2010 article by Marianna Fotaki of the Manchester Business School titled, "Why Do Public Policies Fail So Often?" (Fotaki 2010). It is also the central message in a 2014 book titled "Why Government Fails So Often" by Princeton University's Peter Schuck (Schuck 2014). New York University's Paul Light implies a similar view in his study, "A Cascade of Failures", which suggests that the US federal government may have had some successes in the past but "failures have become so common that they are less of a shock to the public than an expectation" (Light 2014). A 2020 article by Bernard Mueller (Mueller 2020a) offers a similar view. It posits that public policy failure is "ubiquitous" such that, "The public policy fails. It fails often and fails everywhere. From small tasks, such as keeping the streets free of potholes, to massive endeavors such as providing healthcare, it seems that public policy continually comes up short of expectation" (Mueller 2020b).[11]

If true, this claim deserves urgent attention from social and political decision-makers across the world. One might even expect to see these decision-makers actively searching for better ways to solve societal problems (that do not rely on governments and public policy). I do not see evidence of such a shift away from public policy as the preferred way of addressing societal problems, however. The public expenditure data shown earlier suggests precisely the opposite: Societies

[9] Bovens and t'Hart (2016, 653) note, for instance, that "The analysis of policy failures is, by definition, not a neutral endeavor, since policy fiascos are not neutral events."

[10] In referring to the 'myth of "failed" policies', the Independent Institute's Robert Higgs provides an example of the different ways one can think of success and failure, positing that failure actually means success if success involves "channeling resources bilked from the public to politically organized and influential interest groups" (Higgs 1995).

[11] Recent articles with a similar message come Howlett et al. (2015) and Hudson et al. (2019), who speak of the "persistence of policy failure." In a blog post supporting the latter article, Hudson (2019) notes that, "It has always been the case that the likelihood of policy failure is at least as high as policy success. But the currency of modern politics seems to be squarely that of failure — indeed major failure." A recent blog post by Mark Lerner, a fellow at Harvard's Belfer Center, offers similar views: "Government tech projects fail by default" (Lerner 2020). The consulting firm PWC offers data that 60 to 90 percent of government projects in IT fail (PWC 2017). King and Crewe (2013) refer in the same vein to 'the blunders' of governments and Jeff Sachs penned an article in 2009 (on the USA) where he says that "failures are legion" (Sachs 2009). Adding to the chorus, Andrew Graham of Queen's University noted, in 2019, that "Research from around the world shows a consistent pattern of failures in public policy" (Graham 2019).

seem to be entrusting their governments with more, not fewer, policy responsibilities, and are giving these governments control over more resources to meet these expanded responsibilities. Why, if policies fail 'often', do we see this pattern?

Additionally, I do not see public officials embracing policy failure as a common reality they should address. Instead, most officials I encounter consider policy failure as unacceptable and costly, and something they must avoid. As stated in a recent paper by the Center for Public Impact (2017), "In government, where the spotlight of political and public scrutiny can be particularly bright, and the consequences particularly severe, failure is often considered to be 'not an option'."[12] If the officials responsible for policy work hold that failure in this work is 'not an option', can it also be 'ubiquitous' and happen 'often' as the studies quoted earlier suggest? Why, if policies fail as 'often' as claimed, have public officials working on the policies not become more accustomed to failure (and more responsive to it)? And why, if policy failure is indeed 'ubiquitous', do we not see more upheaval in political and bureaucratic systems (the much feared 'consequences' of failure) but instead observe the growth in governmental structures across the world?

I believe we observe these behaviors because the claim that 'public policy failure happens often' has not been convincingly proved, with evidence that demands general and urgent attention (from social decision makers and policy workers). Advocates of this failure narrative typically base their claims on limited studies—of specific policy types or cases—and try to extrapolate the narrow findings from such to the general population of policy work. Mueller's claim that policy failure is ubiquitous is based, for instance, on a limited three-page description of research about policy work in the areas of microcredit and fishing rights. Light's 2014 study revisits stories about 'catastrophic failures' in the United States, centered on (amongst others) mistakes made in intelligence systems preceding the 2001 terrorist attacks on New York, weaknesses exposed in regulatory regimes during the 2008 financial crisis, and gaps that were exposed in emergency response mechanisms and exacerbated the impact of Hurricane Katrina in 2005. Light implies that these examples are

[12] I often hear about 'fear of failure' *often*. In a 2014 essay on the failures associated with the USA's Healthcare.gov policy reforms, Ashley Trim (2014) notes that the fear of failure was a major reason civil servants did not do all they could for the initiative to work. This fear of failure is written about across the world (Adu and Shenge 2012, Kuehnhanss et al 2017, Rego et al 2010, Sihag 2016). In an opinion piece on Canadian government, Young (2014) notes that "when government employees are so scared of failure, they are less likely to try risky things."

representative of systemic policy failures across government, even though their number is limited and they are quite peculiar (mostly exposed by some kind of crisis, for instance). Light does note that the federal government has also had "underappreciated successes" but says most of these happened in the past and should not "obscure the recent cascade of failures" (Light 2014).

Unfortunately, however, he and others do not offer broad-based analysis on how common the 'underappreciated successes' are, if 'catastrophic failures' outnumber such successes, or if most policies actually produce more ambiguous results fitting somewhere between 'underappreciated successes' and 'catastrophic failures'. I thus struggle to find convincing evidence in his work—or others—supporting the narrative that public policy failure is happening in some kind of 'cascade' or that this failure is 'ubiquitous' or that public policies fail 'often'.[13] I also find limited evidence in primary data sources one might look to for proof. Government audit reports, for instance, reflect only on the compliance of spending behavior to legislated budgets (a small part of any discussion on policy success or failure)[14] and public sector performance data is usually produced for some policies and not others (and is seldom comparable across different policy types).[15] Citizen surveys and polls—on satisfaction with government services or trust in government—are also of limited value. They are based on imprecise perceptions, can be difficult to interpret, and offer inconsistent messages.[16] For instance, OCED polls on 'trust in government' show that 45 percent of citizens trusted their government in 2019 (suggesting major discontent with policy and perhaps even high levels of policy failure), but the same polls indicated that 56, 66, 69 and 77 percent of citizens were happy with judicial, education, healthcare and policing systems (and policies) (OECD 2019). Such data hardly suggests that public policies fail often, or that failure is ubiquitous.

[13] There are important texts that tackle the issue in balanced ways—recognizing the diversity of policy results—including Bovens et al (2002) and Compton and t'Hart (2019), but even these studies tend to report on specific case or small sample studies. There are few studies that do analyze bigger samples (like Andrews et al. (2006), which examines performance in 120 local English authorities) but most try and explain failure instead of reflecting on its incidence or the nature of the failure.

[14] For references on this limitation, consider Andrews (2001) and Bourn (2004).

[15] See, for example, the dashboard of Australian performance (https://dashboard.gov.au). It is incomplete and difficult to work with when trying to develop a view on overall government performance. Consider, also, cases on the limits to performance budgeting initiatives intended to connect funding to results in diverse contexts like Florida (Vanlandingham et al. 2005), Bolivia (Montes and Andrews 2005), Ghana (Roberts and Andrews 2005), Tanzania (Ronsholt et al. 2005) and Virginia (Hill and Andrews 2005).

[16] For views on how citizen satisfaction and policy success interact, see Nye et al. (1997) and Van de Walle (2018).

This lack of evidence about frequent failure is, I believe, why we do not see social decision-makers questioning the value of their ever-growing list of policy efforts more often—and why public policy officials designing and implementing policies maintain the narrative that 'failure is unacceptable' (something they would not be able to do if data showed uncontested failure as commonplace).

So … How Regularly Do Policies Fail or Succeed?

Available data on public policy performance thus offers an (at best) unclear story about policy results. A clearer story is only possible with large-sample research, where one examines data showing the actual results of a broad set of policy engagements to determine how often policies succeed or fail. The World Bank's Independent Evaluation Group (IEG) provides such a dataset. The World Bank is a multilateral entity that supports policy initiatives in over 150 governments, in most sectors common to government work—from agriculture to energy and the environment, health care, private sector development, urban organization, transport and water. Policy support in these and other areas comes through multi-year projects that the World Bank helps design, finance and monitor—but are the ultimate responsibility of governments involved. These governments embrace the projects as activities in their own policy portfolios, focused on addressing societal problems they have recognized as needing to be contained or solved.[17]

As with most policy work, these operations start with identifying a problem and designing a multi-year program of action to address the problem. A rural community may be struggling to grow economically, for instance, because of poor linkages to other markets, which results in policy experts designing a transport project to connect community members to new opportunities. The program is then approved and financed, and implementing agents begin with specified activities, hoping to combine such into new outputs and outcomes. In the example, contractors may be hired to design transport hubs and roads (outputs) that they hope to turn into transportation systems that citizens and businesses will use to engage with new markets (where actual use of the hubs

[17] I do not believe that all problems are solve-able, agreeing instead with (Hanberger 2003, 261) that "public policy deals with partly unsolvable or ongoing problems as well as solvable problems." This which means that policy success cannot always be assessed by asking if problems have been solved. Continuous containment can also be success.

and roads and increased levels of citizen and business contacts might be seen as outcomes). Ultimately, the goal is for these interventions to contain or solve the problem permanently (enjoying sustained impact). In the example, we might hope to see new levels of mobility and contact yielding new economic opportunities for the rural community, manifest finally in sustained and expanded economic activity and growth.

The World Bank's IEG evaluates hundreds of operations like this every year, using the same methods in every case, and makes the evaluation ratings and documents available to the public. I examined the data for 999 projects evaluated between 2016 and 2020.[18] The sample is limited to projects evaluated after 2015 to capture apparent changes in evaluation methods introduced in that year.[19] The 999 projects still cover an impressive set of public policy engagements, however, accounting for over $60 billion in funding allocated to policy activities in over 100 low, middle and high-income countries in all the world's regions. I focused my analysis on two of the ratings offered for these projects; 'outcome' and 'risk to development outcome (RDO)'.[20]

The distribution of policy 'outcome' ratings

The 'outcome' rating is arguably the focal metric in World Bank evaluations.[21] It captures the success of policy activities undertaken during the life of a project (which can range from one year

[18] I accessed the data in July 2020 from the IEG site (https://ieg.worldbankgroup.org). The IEG still provides this dataset on the site, but the data accessed through the downloads have changed. In particular, the dataset no longer shows risk to development outcomes scores (RDO scores referenced in this paper).

[19] The changes are ones I noticed in doing analysis of IEG scores for 'risk to development outcomes' indicators. Prior to 2016, there was a clear ranking with 4 possible ratings: 'negligible to low', 'moderate', 'significant', and 'high'. During 2016, evaluations started to register a 'low' rating as separate from 'negligible to low' (which was sometimes only 'negligible'. I assessed the explanations behind thirty 'negligible' and 'low' ratings and it became obvious to me that a fifth rating—'low'—had been added (between 'negligible to low' and 'moderate'. Given this, I decided only to use projects submitted after this change seemed to occur.

[20] As noted in footnote 16, the IEG database now available does not include the RDO rating. It has never showcased the full impressive and comprehensive methods and ratings used in evaluations documents, which go beyond 'outcomes' and 'risks to development objectives'. Other ratings include quality of entry, quality of supervision, overall Bank performance, government performance, implementing agency performance, and quality of monitoring and evaluation. I do not focus on these ratings here because they are commonly referenced as variables that explain results, where the outcome and risk to development outcome ratings are seen as the main results metrics. Put another way, the outcome and risk to development outcome ratings are usually referenced as measures of success and failure while the other ratings are used to measure factors influencing success or failure.

[21] This is the metric the World Bank uses to reflect on its own success (Behrens at al. 2018). It is the metric used in the annual Results and Performance of the World Bank Group (Rap) report (the 2020, version of which is available at:

to over ten years), based on answers to three questions that probe what the IEG calls efficacy, relevance and efficiency.[22] I paraphrase these as follows: Did the project deliver promised outputs and outcomes? Were these deliverables appropriate responses to relevant problems in the specific context? Was the work performed in an efficient manner (on time and on budget and with a reasonable economic return)? Based on answers to such questions, projects are rated as highly unsatisfactory, unsatisfactory, moderately unsatisfactory, moderately satisfactory, satisfactory or highly satisfactory. This is equivalent to ranking success on a six-point scale, where scores of 1 and 2 relate to clear failure (highly unsatisfactory or unsatisfactory outcomes) and scores of 5 and 6 relate to clear success (satisfactory or highly satisfactory outcomes[23]).

Figure 1 shows that projects in the latter two 'success' categories (satisfactory or highly satisfactory outcomes) account for about 30 percent of all policy results. An example comes from China, where the 2010-2016 Integrated Forestry Development project helped advance policies to expand forest coverage and farmer support for forests (given that the country suffers from low levels of forestation and related problems of land degradation).[24] Through an intervention that was considered appropriate to the problem in its context, the policy work delivered (with substantial efficiency) more forests (increased vegetative cover) than targeted, with appropriate species diversity, contributing to reduced run-off and wind speed, increased crop yields, and demonstrable economic benefit to farmers (including the adoption of new economic species, increased yields and higher incomes). Farmers were also trained, and local cooperatives and associations were strengthened, which contributed to the sense of 'success'.

As noted, this project fits into the 30 percent of policy engagements with 'successful' outcome results shown to the right in Figure 1. The figure also shows that there were not many failures (with highly unsatisfactory and unsatisfactory projects accounting for only 10 percent of the total, illustrated at the left side). One such intervention was in the Democratic Republic of the Congo

https://ieg.worldbankgroup.org/evaluations/rap2020. It is also the indicator most researchers use to analyze and explain World Bank success (see, for examples, Honig 2020, Kilby and Michaelowa 2019, and Winters 2019).

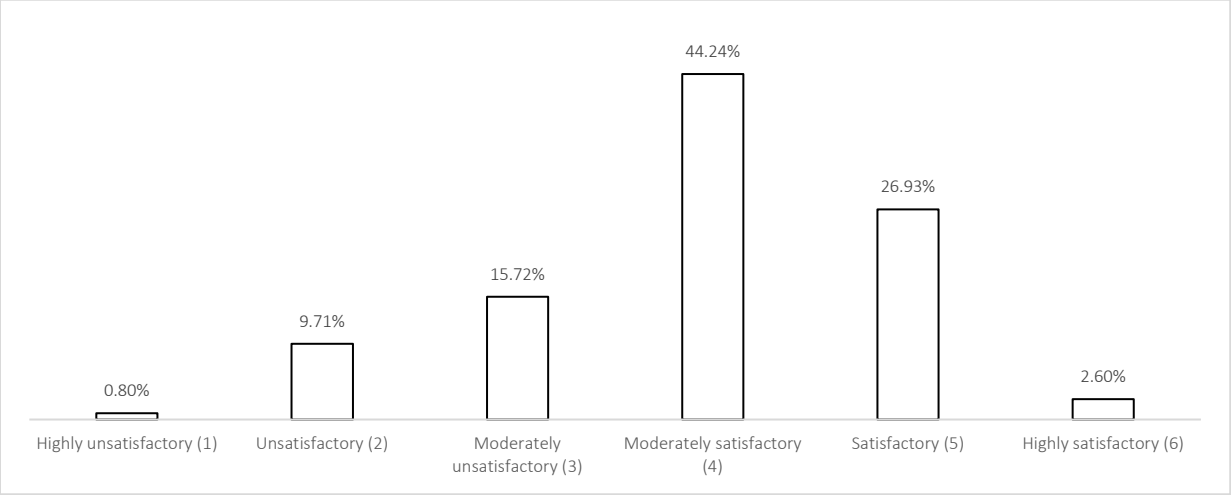
[22] More detail is available at Independent Evaluation Group (2015).

[23] Where a project produced its pre-defined outputs and outcomes efficiently, and these outputs and outcomes are relevant to the context-specific policy concerns. See IEG (2015).

[24] This section draws on the Implementation Completion Report Review (Independent Evaluation Group 2017).

and involved efforts to promote conservation, empower communities living in conservation areas (informing them of their legal rights, potential ways to control illegal logging, and more), build state capability in conservation, and promote deforestation.[25] This project was given an unsatisfactory outcome rating because most proposed activities did not happen and (according to evaluation reports) probably should not have been programmed at all (given weak capacity to implement and monitor the proposed activities, outputs, and outcomes).

Figure 1. The distribution of outcome ratings in a broad sample of World Bank projects



Source: Author's analysis of 'outcome' data in the World Bank IEG project database (accessed in July 2020).

Most policy initiatives in the sample (60 percent) fall into ambiguous 'outcome' categories of moderately unsatisfactory (scoring 3) or moderately satisfactory (scoring 4). Argentina's 2008-2017 rural development project is an example of a category 3 (moderately unsatisfactory) initiative.[26] In keeping with similarly rated projects, the operation essentially produced underwhelmingly mixed results. It focused on a highly relevant problem—the low productivity of small and medium sized agricultural producers—but adopted a strategy that evaluators deemed unclear and unrealistic for the context. Some outputs in this strategy were delivered, including 19

[25] This section draws on the Implementation Completion Report Review (Independent Evaluation Group 2016).
[26] This section draws on the Implementation Completion Report Review (Independent Evaluation Group 2018).

investment plans for provincial departments supporting producers, infrastructure needed to open 3,274 new hectares for production, and significant mileage of rehabilitated irrigation systems. Other outputs were not delivered, however, and the focal outcomes (improved production and sales of small and medium sized entities) were not realized (or were achieved at far lower rates than programmed), resulting in lower than expected economic and financial returns.

Fiji's 2016-2017 Post-Cyclone Winston Emergency project is a good example of the kind of mixed 'outcome' results commonly deemed positive (but not outright successful) in moderately satisfactory initiatives.[27] A policy program considered relevant to the context, focused on financing relief and rehabilitation activities to strengthen the country's resilience to other natural disasters, and providing social protection to the most vulnerable. Many outputs were delivered, including a post-disaster needs assessment, disaster recovery framework (DRF), and evaluation of the disaster-responsive social protection programs. The project also achieved a key outcome, with 96 percent of eligible households and individuals receiving additional financial benefits within the set deadline (exceeding the target of 75 percent). Other outcomes were not achieved, however, reflected most prominently in the partial implementation of the DRF (with only 46 percent of the total budget allocation for DRF activities being spent, and key policy interventions—like school rehabilitation—incomplete at project's end). The policy intervention yielded gains, therefore, but did not meet enough of its outcomes and objectives to be credited as a clear success.

Moderately satisfactory projects (like this crisis response work in Fiji, where a good number of outputs but only some outcomes are delivered, with varying levels of relevance and efficiency) dominate others as the most common 'outcome' of policy work in the sample. In its annual reports, the World Bank combines this ambiguously positive category with the 'satisfactory' and 'highly satisfactory' categories in reporting overall portfolio success (what is sometimes called the MS+ cut off).[28] This means that the organization regularly claims a success rate of over 70 percent in its policy support. This is hardly evidence of ubiquitous failure or of public policies failing 'often'.

[27] This section draws on the Implementation Completion Report Review (Independent Evaluation Group 2019).
[28] This is the criteria reflected in a 2018 IEG press release (Behrens et al. 2018), for instance.

The distribution of policy 'risk to development outcome' ratings

These within-project 'outcome' data do not tell the full story about failure and success in the policy activities undertaken through World Bank projects (or to a generalized policy population), however. The 'outcome' ratings focus most on whether (and how) the policy initiative achieved its "immediate objectives"—within the prescribed short or medium-term project timeframe—not if these immediate objectives actually help to achieve longer run "overarching development goals" or solve or contain the societal problems that warranted a policy response in the first place.[29]

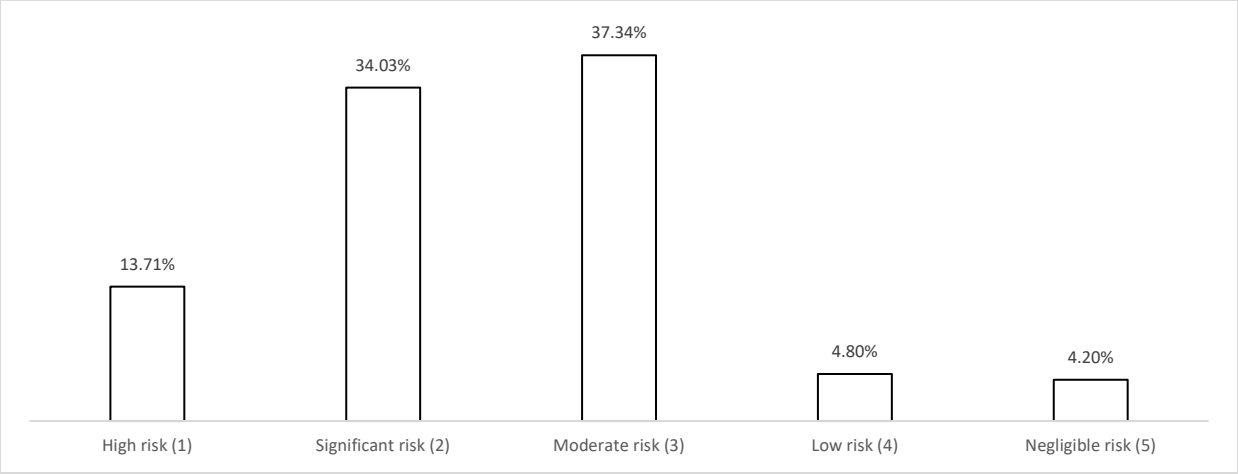
The World Bank IEG does assess the likelihood of this longer-run impact, however, through its rating of the 'Risk to Development Outcome' (RDO).[30] This measure is described as "the risk, at the time of evaluation, that development outcomes (or expected outcomes) will not be maintained (or realized)" (Independent Evaluation Group 2015, p.16). This is not a measure of the actual sustained impact of a policy effort but captures how confident policy experts are—at project completion—that outputs and outcomes will be sustained and impact will be achieved (such that the policy activities actually contain or solve the problems they are intended to address). Different levels of confidence result in projects being assessed into five categories of 'risk to development outcome':[31] High risk (scored as 1) and significant risk (scored as 2) denote major concerns that a policy will not have the sustained impact needed to effectively address focal problems or achieve high-level objectives. An example is Ukraine's 2009-2014 Roads and Safety Improvement engagement.[32] This policy outcome was considered moderately satisfactory (on its 'outcome' rating) because it fostered the rehabilitation of a major highway (the project's central output). At closing, however, evaluators were pessimistic about future maintenance of the road, which raised significant questions about the policy's sustainability and impact. Similarly, policy work yielded a satisfactory 'outcome' rating in Egypt between 2008 and 2015, by expanding "household gas

[29] I draw from Jurgen Blum's discussion of World Bank outcome data—especially Blum's draft paper (2014, p. 28), which was different to the final World Bank working paper he authored on the topic.
[30] As footnote 18 notes, RDO scores are no longer provided in the IEG database. They are, however, key indicators in evaluation documents.
[31] Independent Evaluation Group (2015, p.16) lists four potential ratings. As discussed in footnote 18, I noticed that evaluations used five ratings since 2016, however, splitting 'negligible to low' into 'negligible'/'negligible to low' and 'low'. I work with five ratings given this observation, though it is unclear how formal this convention has become.
[32] This section draws on the Implementation Completion Report Review (Independent Evaluation Group 2015a).

connections in the Greater Cairo Area and elsewhere in the country" but was deemed a significant 'risk to development outcome' because of financial difficulties in the natural gas sector, which made the service provider financially unviable and the continuation of services unlikely.[33]

The message in such high and significant 'risk to development outcome' ratings is that these policies are expected to fail in the long-run and will not, therefore, ultimately solve the problems they were meant to address. This narrative is associated with 48 percent of all projects in the sample (shown in the two left-hand bars in Figure 2). One finds more positive stories at the other, right-hand, extreme, in the figure below.

Figure 2. The distribution of risk to development outcome ratings in 999 World Bank projects



Source: Author's analysis of 'risk to development outcomes' (RDO) data in the World Bank IEG project database (accessed in July 2020). The different categories do not add up to 100 percent because a small number of projects do not record a value for RDO (choosing, instead, to offer a 'not applicable' response).

Policy initiatives shown at the right side of the figure were assessed as low or negligible 'risk to development outcome' (scoring 4 or 5), given a view that their activities will be sustained in the long run and foster the achievement of more advanced impacts. In other words, these policy activities are expected to successfully and sustainably resolve or contain the societal problems

[33] This section draws on the Implementation Completion Report Review (Independent Evaluation Group 2016a).

that warranted a policy response in the first place (or at least contribute significantly to such result). Mexico's 2010 to 2014 initiative to introduce energy efficient technologies in households is an example.[34] Policy activities and outputs included studies on potential energy innovations, the delivery of millions of new energy efficient light bulbs and appliances, research into and marketing of new opportunities for energy efficient innovation, and training in the distribution and use of new technology. Outcomes included actual, measured, improvements in household energy efficiency. At the end of the project's four years, evaluators saw negligible risk to sustained and even growing impact from this policy work because government had already expanded the appliance replacement and lightning programs, introduced an effective monitoring program to measure progress (based on studies done in the project), identified new priority areas for energy savings, and budgeted the funds this work would require.

Unfortunately, only 9 percent of the policy engagements in this sample garnered such positive 'risk to development outcome' ratings, suggesting that sustained and improved long-run impact is only expected with confidence from one-tenth of all projects. Given this, it appears that policies are expected to fail more often than they succeed in the long run (with significant to high risk engagements outnumbering low and negligible risk projects by a factor of more than five).

This view does not change when recognizing that 37 percent of policy engagements were rated as having moderate risk (scored as an RDO of 3). The common-ness of this rating suggests that it is difficult to assess long run risk to development outcome in many projects (given ambiguity or uncertainty in looking to future conditions of policy work). If these moderate risk projects are added to the projects with low or negligible 'risk to development outcome' ratings, we still have a smaller portion of the sample (46 percent) than that made up by the projects at significant to high risk of long run failure (48 percent). If one adds the projects rated as moderate to those at significant and high risk, one starts worrying about long-run failure in 85 percent of all policy work. Ultimately, therefore, the concern over long-term failure affects 48 to 85 percent of policy work—surely evidence that policies fail 'often' (or at least that they are at risk of such failure).

[34] This section draws on the Implementation Completion Report Review (Independent Evaluation Group 2015b).

A programmatic success distribution, and 'successful failures'

This distribution of 'risk to development outcome' scores indicates a higher likelihood of policy failure than suggested in the earlier 'outcome' ratings distribution (where failure was uncommon, associated with only 26 percent of interventions rated as moderately unsatisfactory or worse). As such, the two sets of data—and the perspectives they offer on how regularly policies succeed and fail—could be interpreted as conflicting. I do not share this interpretation, however, but rather see the data from both ratings combining into one story about how often World Bank supported policies achieve different levels of what the literature calls programmatic success. This story also points to a common programmatic policy result, which I call 'successful failure'.

Programmatic (or program) success is one of various oft-referenced dimensions of public policy success (with other dimensions including policy and process success, and endurance).[35] The programmatic dimension captures whether public policy interventions produce "the results desired" (McConnell 2010, p.353) or "really have an impact on the issues they are supposed to tackle" (Compton and t'Hart 2019, p.1). Such success is commonly assessed by asking about the quality of the work program (did it address relevant problems in a contextually relevant way?), its demonstrated results (did it deliver and achieve what was promised, in the short, medium, and long run?), and its efficiency (was it cost and time efficient?). The 'programmatic success' focus of IEG evaluations is reflected in how it captures such things (relevance, efficacy, efficiency, and the likelihood of sustained impact) in 'outcome' and 'risk to development outcome' ratings.[36]

[35] Literature suggests assessing policy success on multiple dimensions, emphasizing programmatic, political and process performance, legitimacy, and sustainability (Bovens et al. 2001, Boyne 2003, 2004, Wallner 2008, Howlett 2012, Rainey 2003, Marsh and McConnell 2010, McConnell 2010a, 2010b, and Compton and t'Hart 2019). Process and political dimensions relate to how policy work impacts political support and governing legitimacy, given what work is done and how it is chosen, how work is done (if means are considered legitimate) and who pays and benefits (given distributed costs and gains). When combined, such dimensions contribute to a view that a policy succeeds when engagements achieve programmatic results, follow processes that stakeholders consider appropriate, retain political support, and endure, or "if it achieves the goals that proponents set out to achieve and attracts no criticism of any significance and/or support is virtually universal" (Compton and t'Hart 2019, pp.4-5; McConnell 2010, p. 348).

[36] One might argue that the World Bank also embeds a concern about process into its outcomes rating (when asking about relevance of objectives and design) and that the 'risk to development outcome' rating reflects what Compton and t'Hart (2019) call sustainability. I see 'risk to development outcome' ratings reflecting long-term programmatic success (combining Compton and t'Hart's idea of sustainability into a long-run view of programmatic success).

Programmatic success can be assessed at various 'levels', given the achievement of progressive performance milestones in a policy program.[37] Development organizations like the World Bank frequently identify such levels with reference to concepts like inputs, outputs, outcomes, and impacts (Klingebiel and Janus 2014, World Bank 2007, 2020, Wolman 1981).[38] The lowest level of success recognized in such entities typically involves mobilizing programmed inputs (like finances) to facilitate activities (like road building) and the production of outputs (like new roads).[39] The next level of success is reached when combinations of new outputs yield initial outcomes (like better access to transport for farmers) and even some intermediate outcomes (like improved farmer connections to markets).[40] Higher levels of success are achieved when significant intermediate outcomes are forthcoming, sustained and improved, and promote measurable impact—such that focal problems are solved or contained (where, in my example, farmers are sustainably connected to new markets where they sell their goods, leading to continuing periods of economic growth).[41]

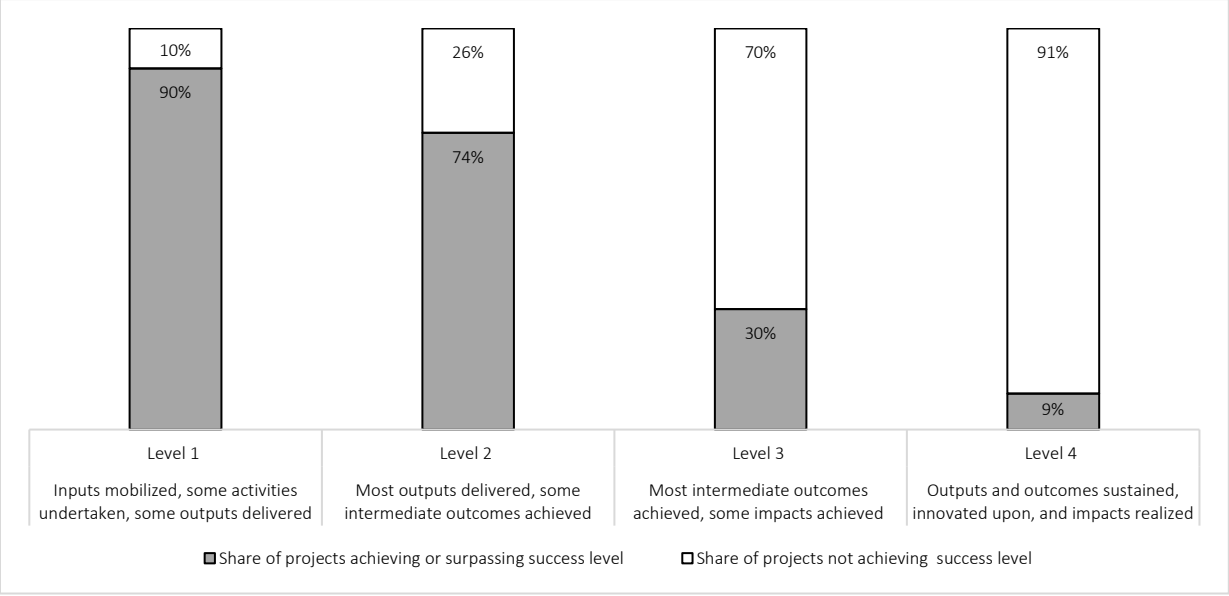
Efficiency and program quality matter at all levels of this programmatic success, with cost and time overruns or inappropriate designs at any point in an intervention impeding future milestone achievement. For instance, delayed, over-budget and low-quality delivery of parts of a road project will threaten efforts to build a full road network and ensure this gives farmers broad market access.

This line of thinking informs a 4-level programmatic success framework shown in Figure 3.[42] The figure also illustrates how regularly I see policy interventions achieving (or surpassing) different

[37] I take the word 'level' from Paul Bannerman's application of this kind of thinking to develop a model of project management success (Bannerman 2008, p.6). McConnell (2010) might call these 'degrees' of success.
[38] I draw on IEG 'outcome levels' here (Independent Evaluation Group 2020). These levels differentiate between the different products and results (or outcomes) emerging from World Bank projects.
[39] As defined in World Bank (2007), "Inputs are the financial, human, and other resources mobilized to support activities undertaken by a project." "An Activity is an action taken or work performed by which inputs are converted into specific outputs [like] providing training, constructing a facility, etc." "Outputs are supply-side deliverables, including events, products, capital goods or services that result from a[n] intervention (e.g., construction of a school)."
[40] As defined in World Bank (2007), "A Project Outcome is the uptake, adoption or use of project outputs by the project beneficiaries" and "An Intermediate Outcome specifies a result proximate to an intended final outcome, but likely more measurable and achievable in the lifetime of a project to an intended final outcome."
[41] As defined in World Bank (2007), "Higher Level Outcomes are the long-term goals near or at the top of the Results Chain, e.g., "Reduce Child Mortality. Achievement of higher-level outcomes is generally beyond the scope of a project. Nevertheless, it is important that a project identify the higher level outcome it seeks to influence, i.e., the country development goal … that the project together with other interventions (Bank and other donors) expects to influence."
[42] As in footnote 38, I draw on IEG's different 'outcome levels' here (Independent Evaluation Group 2020).

levels of this success (based on my analysis of IEG data[43]). For instance, I recognize 90 percent of the operations as reaching level-1 success (where some outputs were delivered on time and on budget).[44] I show an impressive share of projects—74 percent—achieving at least level-2 success (where most outputs and some intermediate outcomes were delivered and achieved, on time and on budget, satisfying the World Bank's key MS+ success reporting criteria).[45] Only 30 percent of the projects attained level-3 success, however, where most intermediate outcomes and some impacts were achieved (on time and on budget).[46] I credit even fewer policy engagements (9 percent) as achieving level-4 success, where most outcomes were achieved and expected to be sustained and improved upon and to yield necessary impact (contributing to solving or containing the problems that inspired and warranted the policy work).[47]

Figure 3. How regularly World Bank projects achieved different levels of programmatic success



Source: Author's original work, based on analysis of IEG project evaluations.

[43] It must be noted that this is my interpretation of the IEG ratings.
[44] I do not credit the 10 percent of highly unsatisfactory and unsatisfactory projects with reaching this level.
[45] This excludes the 26 percent with highly unsatisfactory, unsatisfactory, and moderately unsatisfactory ratings.
[46] These are the operations with satisfactory and highly satisfactory outcomes.
[47] These are the projects that produced 'satisfactory' or 'highly satisfactory' outcomes and where the risk to development outcomes was considered 'negligible' or 'low'.

As presented, Figure 3 merges the two sets of data shown in earlier figures, telling a singular story about how regularly World Bank supported policy interventions achieve or surpass different levels of programmatic success. This story suggests that outright level-1 policy failure and outright Level-4 policy success are rare, accounting (together) for only 19 percent of policy engagements. Most policies delivered more ambiguous and mixed mid-level success, with the vast majority reaching only mid-levels 2 and 3. These are the 'successful failures' of policy work—operations that delivered some to most programmed outputs and outcomes and even some short-term impacts, but where no one expected such gains to sustain or yield higher-level, long-run impacts needed to solve or contain the societal problems they were intended to address. Consider stylized examples: A health policy initiative could mobilize medical supplies to treat childhood illnesses that no one expects will be used because of weak distribution systems, limiting any gains in child health outcomes; An energy project could deliver a new power station that produces less electricity than planned because of weak post-project management, generating limited economic benefits; A rural development policy intervention might build roads that increase farmer market access for only a short period because of poor maintenance, with limited long-run economic impact; An education project may buy millions of new textbooks for schools, but these may not be delivered efficiently or used properly because of ineffective management systems in the education sector, leading to marginal improvements in learning.

Is Successful Failure Endemic to Public Policy Work?

I expect these examples will resonate broadly with readers, given that governments regularly engage with these kinds of public policy challenges—tackling an ever-expanding set of societal problems their citizens cannot address on their own. Citizens increasingly depend on these public policy initiatives succeeding. Without such success, childhood health outcomes will remain poor in many contexts, electricity access will not be sufficient to facilitate basic activities, farmers will lack access to markets and the economic growth opportunities these markets offer, future generations will learn less than is required to make and take opportunities needed for progress and development, and more. Unfortunately, a prominent view holds that governments 'often' fail

when doing this kind of work. This view would be bad news for citizens but tends to lack a strong evidentiary backing, which makes it hard to know if policies really do fail often or not. The lack of evidence is why I believe we do not see citizens withdrawing their support for 'failed' public policy organizations and why public policy makers do not seem to recognize regular failure (and even hold that failure is actually 'unacceptable' and "not an option" (Center for Public Impact (2017)).

The study described in this paper is a modest attempt to provide large-sample evidence of how regularly policies really succeed or fail. Its findings do not support the prominent failure narrative, with few of the 999 policy initiatives producing what could be called outright failure. Rather, most initiatives yielded an ambiguous result I call 'successful failure'. This result is achieved when a policy delivers low-level short-run outputs and outcomes that are not expected to be sustained or translated into the higher-level, long-run impacts needed to address societal problems.

I posit that such 'successful failure' is endemic to public policy work. It characterized results in up to 65 percent of the policy interventions I analyzed,[48] given policy evaluation data that I believe are as reliable as one is likely to find and reflecting a broad spectrum of policy interventions in over 100 low, middle, and high-income countries.[49] The share of 'successful failures' differs across regions and policy types but is always high and greater than the combined likelihood of registering

[48] These were the projects that achieved programmatic success levels 2 or 3 but did not reach level 4. In the IEG ratings, they scored moderately satisfactory or higher 'outcome' ratings but did not score negligible or low 'risk to development outcomes' ratings.

[49] To address the many data challenges involved in this kind of study, I focus on only one dimension of policy success—the programmatic dimension—and draw data from a highly respected policy evaluation entity, the World Bank's IEG. The IEG has detailed conceptual descriptions and evaluation guidelines that foster indicator validity (ensuring the concepts being rated are clear and consistent) and employs stringent quality control measures to promote score reliability (including review processes) (IEG 2017a). Considering these descriptions and measures, I believe the IEG ratings used in this study—assessing 'outcome' and 'risk to development outcome' results—offer sufficiently valid and reliable measures of different levels of programmatic success to facilitate the analysis undertaken. I am not the first one to use such scores in academic research. Researchers commonly reference the 'outcome' rating, endorsing such as a valid result metric based on actual evidence of results achieved in a project (Andrews 2018, Honig 2018, 2020). The 'risk to development outcome' metric is not as commonly referenced and arguably more subjective—given that it captures evaluators' views on the risks associated with future success—but it, too, is based on specific methods and is always accompanied with a clear scoring rationale. For instance, IEG evaluation guidelines (2017, pp.62-63) provide detail on what the "risk to development outcome" is: "[T]he risk, at the time of evaluation, that development outcomes (or expected outcomes) will not be maintained (or realized). This refers to outcomes that have actually been achieved (or are expected to be achieved)." The document also offers "Guidance" on how to assess it, noting that, "The risk to development outcome has two dimensions: (i) The likelihood that changes may take place that are detrimental to the ultimate achievement of the project's development outcome, (ii) The impact of some or all of these changes on the project's development outcomes."

an outright policy success or failure.[50] The narrative of 'successful failures' is common, and reflected in my four introductory stories from the field, which show that the result does not necessarily reflect World Bank involvement (as some might theorize, given the sample): Brazil's government did engage the World Bank to successfully deliver educational outputs and intermediate outcomes (like school access) that have yet to yield needed gains in education quality; In contrast, the French government did not need World Bank help to succeed in delivering laws and IT systems that have failed (for two decades) to control health care costs (the problem they were meant to address); South Africa's government also did not enjoy the World Bank's help when generating economic growth plans that failed to halt its growing unemployment problem; Similarly, Australia's government worked alone to successfully craft the laws, reports and agencies that have repeatedly failed to address emissions and limit the pace and impact of climate change.

Literature helps explain why policy actors and their organizational, political, and civic systems produce such 'successful failure'.[51] To start, studies note that actors involved in policy design and implementation are often risk averse and fearful of being blamed for failure. Their fear is amplified when policy demands are uncertain, given "imperfect knowledge" or an "inability to … predict the future" (Nair and Howlett 2017, pp.103-104). They manage such uncertainty with credit claiming and blame-avoidance strategies[52] that involve associating with certain, controllable engagements and "distancing" from uncertain, hard-to-control work (Pollitt and Bouckaert 2004). Such behavior causes them to avoid innovating in ways needed to address new societal problems (which, being uncertain and risky, inspires distancing). This behavior also generates a myopic approach to policy work, where actors commit to what is predictable in the near-term but distance themselves from uncertain, risky challenges that are key to achieving high-level, longer-run impact.[53]

[50] The share of projects I would classify as 'successful failures' was 61 percent in Sub-Saharan Africa, 68 percent in East Asia and the Pacific, 74 percent in Europe and Central Asia, 61 percent in Latin America and the Caribbean, 65 percent in the Middle East and North Africa, and 67 percent in South Asia. The share was 74 percent for agriculture projects, 63 percent for projects in the education sector, 65 percent for energy and mining operations, 57 percent for health interventions, 73 percent for projects on urban development, and 57 percent for water policies.

[51] Especially studies on how risk aversion, fear of failure, and uncertainty affect policy work (Hansson 2018, Howlett 2012, 2014, Leong and Howlett 2017, Nair and Howlett 2017, Pollitt and Bouckaert 2004, Weaver 1986).

[52] Fostering self-promotion and self-preservation (maximizing rewards from success and limiting blame for failure).

[53] Nair and Howlett (2017, 103-104) explain that 'policy myopia' arises because of "the difficulty of seeing far enough into the future to discern its general shape and contour in enough detail to be able to properly anticipate and plan in

Literature suggests that public organizations, citizens, and politicians often reinforce such behavior, contributing to make 'successful failures' endemic. Organizational systems frequently reward policy design and implementation actors who successfully deliver safe short-run products, for instance, even if these products do not yield further gains—and penalize actors who fail in the short run trying to achieve more uncertain, higher-level, longer run impacts.[54] Public organizations do this because the myopic citizens and politicians on whom they depend (for support and legitimacy) demand evidence of real time (or foreseeable) results. This is what authors call the "policy bias toward the present" which sees citizens and politicians "weighing short-term consequences" (and even promises) "more heavily than long-term outcomes" (Jacobs and Matthews 2012, p.903). Such bias results in the "political import of longer-term effects" being lower than that for short-term deliverables (even if those deliverables are repeatedly failing to generate important and even vital long-term impacts) (Jacobs 2016, p.437).

Such reasoning helps explain why 'successful failures' are so common in the sample of 999 policy initiatives studied here and strengthens my view that this result is endemic to public policy work. It also suggests that we should worry more about 'successful failure' than we do about outright policy failure. The latter has some chance of attracting a corrective response that limits the time and resource costs of pursuing faulty policy work. In contrast, 'successful failures' are more pernicious—doing harm in gradual and subtle ways—and thus unlikely to attract the kind of civic, political or administrative attention needed to mobilize corrective action. Such results are most likely to generate constant streams of policy activities that absorb time and resources delivering outputs and outcomes that never solve the problems society is struggling with.

---

the present." They offer further that fear about "uncertainties" associated with the future and assumptions of "a greater level of certainty in the short-term" feed 'policy myopia', suggesting this is "a major cause of policy failure."

[54] This is arguably what the World Bank does—and is transparent about doing—when it communicates results based on 'outcome' ratings only (without reference to other indicators like the 'risk to development outcomes' ratings). The 'outcome' ratings capture project performance "within the control of the Bank" (Behrens et al. 2018, World Bank 2020) and incentivize agents to focus on low-level deliverables that they can control and will assure at least a moderately satisfactory score (meeting the MS+ standard). This stifles innovation and a longer-run focus on ensuring sustainability and impact of work products (Honig 2018, 2020, World Bank 2020).

References

Adam, C., Hurka, S., Knill, C. and Steinebach, Y., 2019. Policy accumulation and the democratic responsiveness trap. Cambridge University Press.

Adu, R. and Shenge, N.A., 2012. Organizational culture and psychological factors as predictors of indulgence in procrastination among civil servants in Oyo State, Nigeria.

Andrews, M., 2001. Adjusting external audits to facilitate results-oriented government. International Journal of Government Auditing, 28(2), p.10.

Andrews, M., 2018. Public Policy Failure: "How Often?" and “What is Failure, Anyway?": A Study of World Bank Project Performance. Center for International Development at Harvard University.

Andrews, R., Boyne, G.A. and Enticott, G., 2006. Performance failure in the public sector: misfortune or mismanagement?. Public Management Review, 8(2), pp.273-296.

Aron, J., Kahn, B. and Kingdon, G. eds., 2009. South African economic policy under democracy. Oxford University Press.

Behrens, J, Siy, A.M, and Langley, W. 2018. Are World Bank Group Projects Getting Results? https://ieg.worldbankgroup.org/blog/are-world-bank-group-projects-getting-results

Besley, T. and Case, A., 2003. Political institutions and policy choices: evidence from the United States. Journal of Economic Literature, 41(1), pp.7-73.

Blum, J. R., 2014. What factors predict how public sector projects perform? A review of the World Bank's public sector management portfolio. Washington, D.C.: The World Bank.

Bourn, J. 2008. Public sector auditing: Is it value for money? John Wiley & Sons.

Bovens, M.A., T'Hart, P. and Peters, B.G. eds., 2002. Success and failure in public governance: A comparative analysis. Edward Elgar Publishing.

Bovens, M. and T’Hart, P., 2016. Revisiting the study of policy failures. Journal of European Public Policy, 23(5), pp.653-666.

Burnel, P., 2018. The introduction of electronic medical records in France: More progress during the second attempt. *Health Policy*, *122*(9), pp.937-940.

Center for Public Impact. 2017. Why failing isn't final: A new taxonomy of failure in government. September 14, 2017. https://www.centreforpublicimpact.org/insights/failurenotfinal

Compton, M. and Hart, P.T. eds., 2019. Great policy successes. Oxford University Press.

Crowley, K., 2017. Up and down with climate politics 2013–2016: the repeal of carbon pricing in Australia. Wiley Interdisciplinary Reviews: Climate Change, 8(3), p.e458.

Crowley, K., 2021. Fighting the future: The politics of climate policy failure in Australia (2015–2020). Wiley Interdisciplinary Reviews: Climate Change, 12(5), p.e725.

Fotaki, M., 2010. Why do public policies fail so often? Exploring health policy-making as an imaginary and symbolic construction. *Organization*, *17*(6), pp.703-720.

Grady, A., 2012. Electronic Health Records How the United States Can Learn From the French Dossier Medical Personal. Law school diss., University of Wisconsin Law School. http://hosted.law.wisc.edu/wordpress/wilj/files/2013/01/Grady.pdf.

Graeve, C., 2020. e-Health in France: Spotlight on the National Healthcare Digitalization Strategy. Health Advances Blog, March 24, 2020. https://healthadvancesblog.com/2020/03/24/e-health-in-france/

Graham, A. 2019. Why governments are so bad at implementing public projects. The Conversation. February 13, 2019.

Haddad, M.A., Freguglia, R. and Gomes, C., 2017. Public spending and quality of education in Brazil. The Journal of Development Studies, 53(10), pp.1679-1696.

Hanberger, A., 2003. Public policy and legitimacy: A historical policy analysis of the interplay of public policy and legitimacy. Policy Sciences, 36(3), pp.257-278.

Hansson, S., 2018. The discursive micro-politics of blame avoidance: Unpacking the language of government blame games. Policy Sciences, 51(4), pp.545-564.

Hausmann, R. and Andrews, M., 2009. Why we still believe exports for jobs will lead to shared growth: A response to Fine's" Harvard Group Shores up Shoddy Governance". Transformation: Critical Perspectives on Southern Africa, 69(1), pp.31-65.

Higgs, R. 1995. The Myth of "Failed" Policies. Independent Institute Commentary, March 1, 1995. https://www.independent.org/news/article.asp?id=123

Hill, H. and Andrews, M., 2005. Reforming budget ritual and budget practice: the case of performance management implementation in Virginia. International Journal of Public Administration, 28(3-4), pp.255-272.

Hill, K.Q., Leighley, J.. and Hinton-Anderson, A. 1995. Lower-Class Mobilization and Policy Linkage in the U.S. States, American Journal of Political Science, 39, 75-86.

Honig, Dan. 2018. Navigation by Judgment: Why and When Top-Down Management of Foreign Aid Doesn't Work. Oxford: Oxford University Press.

Honig, D., 2020. Information, power, and location: World Bank staff decentralization and aid project success. Governance, 33(4), pp.749-769.

Horn, R., 2002. Improving schools and schooling in Brazil: The Fundescola approach. The World Bank.

Howlett, M., 2012. The lessons of failure: learning and blame avoidance in public policy-making. International Political Science Review, 33(5), pp.539-555.

Howlett, M., 2014. Why are policy innovations rare and so often negative? Blame avoidance and problem denial in climate change policy-making. Global Environmental Change, 29, pp.395-403.

Howlett, M., Ramesh, M. and Wu, X., 2015. Understanding the persistence of policy failures: The role of politics, governance and uncertainty. Public Policy and Administration, 30(3-4), pp.209-220.

Hudson, B. 2019. We need to talk about policy failure — and how to avoid it. Apolitical. January 23, 2019. https://apolitical.co/solution-articles/en/how-to-avoid-policy-failure

Hudson, B., Hunter, D. and Peckham, S., 2019. Policy failure and the policy-implementation gap: can policy support programs help?. Policy design and practice, 2(1), pp.1-14.

Hudson, M., 2019. 'A form of madness': Australian climate and energy policies 2009–2018. Environmental Politics, 28(3), pp.583-589.

Independent Evaluation Group. 2015. Project Performance Ratings Codebook. Washington, D.C.: World Bank. https://ieg.worldbankgroup.org/sites/default/files/Data/reports/ieg-wb-project-performance-ratings- codebook_092015.pdf

Independent Evaluation Group. 2015a. Ukraine Roads and Safety Improvement Project. Implementation Completion Report (ICR) Review. Washington, D.C.: World Bank IEG. https://documents1.worldbank.org/curated/en/369601468302038180/pdf/ICRR-Disclosable-P100580-02-09-2016-1455055191719.pdf

Independent Evaluation Group. 2015b. Mexico Efficient lighting and appliances. Implementation Completion Report (ICR) Review. Washington, D.C.: World Bank IEG.

https://documents1.worldbank.org/curated/en/959221472852941772/pdf/ICRR-Disclosable-P106424-09-02-2016-1472852934950.pdf

Independent Evaluation Group. 2016. Democratic Republic of Congo Forest and Nature Conservation Project. Implementation Completion Report (ICR) Review. Washington, D.C.: World Bank IEG. https://documents1.worldbank.org/curated/en/442031473302248522/pdf/Congo-Democratic-Republic-of-DRC-Forest-and-Nature-Conserv-PFCN.pdf

Independent Evaluation Group. 2016a. Egypt Natural Gas and Connections Project. Implementation Completion Report (ICR) Review. Washington, D.C.: World Bank IEG. https://documents1.worldbank.org/curated/en/245211467997255959/pdf/ICRR14839-P095392-Box394823B-PUBLIC.pdf

Independent Evaluation Group. 2017. China Integrated Forestry Development Project. Implementation Completion Report (ICR) Review. Washington, D.C. : World Bank IEG. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/873801512769773616/china-cn-integrated-forestry-development.

Independent Evaluation Group. 2017a. DRAFT Guidelines for Reviewing World Bank Implementation Completion and Results Reports: A Manual for IEG ICR Reviewers. https://ieg.worldbankgroup.org/sites/default/files/Data/More-on-IEG/icrr_evaluatormanual_august2018.pdf

Independent Evaluation Group. 2018. Argentina PROSAP2 (2nd Provincial Agricultural Development Project. Implementation Completion Report (ICR) Review. Washington, D.C. : https://documents1.worldbank.org/curated/en/957061516292814141/pdf/ICRR-Disclosable-P106684-01-18-2018-1516292791420.pdf

Independent Evaluation Group. 2019. Fiji Post-Cyclone Winston Emergency DPO. Implementation Completion Report (ICR) Review. Washington, D.C. : World Bank IEG. https://documents1.worldbank.org/curated/en/213691549909250766/pdf/Fiji-Fiji-Post-Cyclone-Winston-Emergency-DPO.pdf

IHE. 2018. IHE Implementation Case Study: French Electronic Health Record Program (n.d.). Integrating the Healthcare Enterprise. Retrieved from http://www.ihe.net/uploadedFiles/Content/case_study_france_ehr.pdf.

Jacobs, A.M., 2016. Policy making for the long term in advanced democracies. Annual Review of Political Science, 19, pp.433-454.

Jacobs, A.M. and Matthews, J.S., 2012. Why do citizens discount the future? Public opinion and the timing of policy consequences. British Journal of Political Science, 42(4), pp.903-935.

Kilby, C. and Michaelowa, K., 2019. What influences World Bank project evaluations?. In Lessons on Foreign Aid and Economic Development (pp. 109-150). Palgrave Macmillan, Cham.

King and Crewe. 2013.

Klingebiel, S. and Janus, H., 2014. Results-based aid: potential and limits of an innovative modality in development cooperation. International Development Policy| Revue internationale de politique de développement, (5.2).

Kuehnhanss, C.R., Murdoch, Z., Geys, B. and Heyndels, B., 2017. Identity, threat aversion, and civil servants' policy preferences: Evidence from the European Parliament. Public Administration, 95(4), pp.1009-1025.

Leong, C. and Howlett, M., 2017. On credit and blame: Disentangling the motivations of public policy decision-making behaviour. Policy Sciences, 50(4), pp.599-618.

Lerner, M. 2019. Government tech projects fail by default. It doesn't have to be this way. Harvard Kennedy School Belfer Center Perspectives on Public Policy Purpose Blog Post. October 21, 2020. https://www.belfercenter.org/publication/government-tech-projects-fail-default-it-doesnt-have-be-way

Light, P.C., 2014. A cascade of failures: Why government fails, and how to stop it. Brookings Institution, July, 14.

Lott, John R., Jr. and Kenny, Lawrence W. 1999. Did Women's Suffrage Change the Size and Scope of Government?," Journal of Political Economy, 107(6), 1163-98.

Manard, S., Vergos, N., Tamayo, S. and Fontane, F., 2019. Electronic health record in the era of industry 4.0: the French example. *arXiv preprint arXiv:1907.10322*.

Marsh, D. and McConnell, A., 2010. Towards a framework for establishing policy success. Public administration, 88(2), pp.564-583.

McConnell, A., 2010. Policy success, policy failure and grey areas in-between. Journal of public policy, 30(3), pp.345-362.

Montes, C. and Andrews, M., 2005. Implementing reforms in Bolivia: Too much to handle?. International Journal of Public Administration, 28(3-4), pp.273-290.

Mueller, B., 2020a. Why public policies fail: Policymaking under complexity. EconomiA, 21(2), pp.311-323.

Mueller, B. 2020b. Why public policies fail. FSR Energy and Climate. April 6, 2020. https://medium.com/lights-on-innovation/why-public-policy-fails-719891913f5a

Müller, P. and Slominski, P., 2017. The politics of learning: developing an emissions trading scheme in Australia. Global Environmental Politics, 17(3), pp.51-68.

Naidoo, V. and Maré, A., 2015. Implementing the National Development Plan? Lessons from coordinating grand economic policies in South Africa. Politikon, 42(3), pp.407-427.

Nair, S. and Howlett, M., 2017. Policy myopia as a source of policy failure: Adaptation and policy learning under deep uncertainty. Policy & Politics, 45(1), pp.103-118.

Nye, J., Zelikow, P. & King, D. 1997. Why People Don't Trust Government. Cambridge, M.A. : Harvard University Press.

OECD. 2019. Government at a Glance. https://www.oecd.org/statistics/public-governance-a-matter-of-trust.htm

Ortiz-Ospina, E. and Roser, M. 2020. Government Spending. Our World in Data. https://ourworldindata.org/government-spending

Pollitt, C. and Bouckaert, G., 2004. Public management reform: A comparative analysis. New York: Oxford University Press.

PWC. 2017. Are public projects doomed to fail from the start? https://www.pwc.ch/en/insights/public-sector/are-public-projects-doomed-to-fail.html

Raiser, M., 2018. Brazil can improve education by copying its own successes. Future Development Blog, Brookings. March 6, 2018. https://www.brookings.edu/blog/future-development/2018/03/06/brazil-can-improve-education-by-copying-its-own-successes/

Rego, A., Marques, C., Leal, S., Sousa, F. and Pina e Cunha, M., 2010. Psychological capital and performance of Portuguese civil servants: Exploring neutralizers in the context of an appraisal system. The International Journal of Human Resource Management, 21(9), pp.1531-1552.

Roberts, P., 2006. FEMA after Katrina. Policy Review, 137(15), pp.15-33.

Roberts, J. and Andrews, M., 2005. Something funny happened on the way to reform success: the case of budget reform implementation in Ghana. International Journal of Public Administration, 28(3-4), pp.291-311.

Ronsholt, F.E. and Andrews, M., 2005. "Getting it together"… or Not. An analysis of the early period of Tanzania's move towards adopting performance management systems. International Journal of Public Administration, 28(3-4), pp.313-336.

Sachs, J. 2009. The Failing U.S. Government—The Crisis of Public Management. Scientific American. October 1, 2009.

Schuck, P.H., 2014. Why government fails so often. Princeton University Press.

Schwartzman, S., 2005. Education-oriented social programs in Brazil: the impact of Bolsa Escola. Instituto de Estudos do Trabalho e Sociedade.

Séroussi, B. and Bouaud, J., 2016. Adoption of a nationwide shared medical record in France: Lessons learnt after 5 years of deployment. In *AMIA annual symposium proceedings* (Vol. 2016, p. 1100). American Medical Informatics Association.

Séroussi, B. and Bouaud, J., 2018. The (re)-relaunching of the DMP, the French shared medical record: new features to improve uptake and use. In *Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth* (pp. 256-260). IOS Press.

Séroussi, B. and Bouaud, J., 2020. Update on the DMP, the French Nationally Shared Medical Record: Did We Make It?. *Studies in health technology and informatics*, *270*, pp.698-702.

Sihag, A., 2016. Aspirants to civil servants: Case studies. Indian Journal of Health and Wellbeing, 7(7), p.763.

Thwala, W.D., 2011. Public works programmes as a tool to address unemployment and skills shortages among the youth in South Africa. African Journal of Business Management, 5(15), pp.6011-6020.

Trim, A. 2014. Government bureaucracy's fear of failure doomed healthcare.gov. Zocalo Public Square Essay. January 8, 2014.

Van de Walle, S., 2018. Explaining citizen satisfaction and dissatisfaction with public services. The Palgrave handbook of public administration and management in Europe, pp.227-241.

VanLandingham, G., Wellman, M. and Andrews, M., 2005. Useful, but not a panacea: performance-based program budgeting in Florida. International Journal of Public Administration, 28(3-4), pp.233-253.

Wallner, J., 2008. Legitimacy and public policy: Seeing beyond effectiveness, efficiency, and performance. Policy Studies Journal, 36(3), pp.421-443.

Weaver, R.K., 1986. The politics of blame avoidance. Journal of public policy, 6(4), pp.371-398.

Winters, M.S., 2019. Too many cooks in the kitchen? The division of financing in World Bank projects and project performance. Politics and Governance, 7(2), pp.117-126.

Wolman, H., 1981. The determinants of program success and failure. Journal of Public Policy, 1(4), pp.433-464.

World Bank., 2007. Results Terminology. Paper Prepared by the OPCS Results Secretariat – September 5, 2007. https://thedocs.worldbank.org/en/doc/265271525990134051-0090022018/original/04WorldBankResultsTerminology.pdf

World Bank., 2018. Annual Report. Washington, D.C.: World Bank.

World Bank., 2020. Results and Performance of the World Bank Group (RAP). Washington, D.C.: World Bank.

Young, W. 2018. Fear of failure holds federal government back. iPolitics Opinions, November 14, 2018.