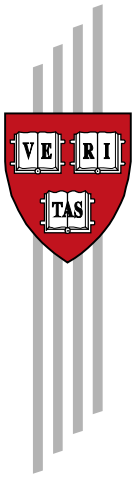
Public Policy Failure: 'How Often?' and 'What is Failure, Anyway'?

A Study of World Bank Project Performance

Matt Andrews

CID Faculty Working Paper No. 344
December 2018



Working Papers

Center for International Development at Harvard University

Public Policy Failure: 'How Often?' and 'What is Failure, Anyway'?

A Study of World Bank Project Performance

Matt Andrews

Harvard Kennedy School

Abstract

Observers claim that public policies fail 'often'. This paper asks, 'how often'? It is an important question, because public policies absorb resources to address major social issues. We should know if policies are proving bad social investments; routinely failing to solve focal problems at high costs. Unfortunately, it is not easy to assess this. Many public policy organizations—governments in particular—do not provide accessible views onto overall success or failure. The World Bank does, however, provide such view—and it supports policy interventions one finds in governments across the world. The paper thus examines World Bank failure rates. It finds that there are different answers to the 'how often' question, depending on responses to a second question, 'what is failure anyway?' In studying both questions, the paper identifies a bias in the World Bank—and probably all organizations adopting rational 'plan and control' policy processes—to measuring 'project and product success' rather than a broader view of success as 'problems are solved with development impact'. This means that policy organizations like the Bank judge success based on whether planned products are delivered through an efficient process; not whether policies solve the problems that warranted intervention in the first place, or whether the policies promoted development outcomes. Is this how citizens would want their public policy organizations to conceptualize success?

Do Public Policies Really Fail 'So Often'?

In 2010, the Manchester Business School's Marianna Fotaki authored an article titled "Why Do Public Policies Fail So Often?" In 2014, Princeton University's Peter Schuck wrote a similarly titled book, "Why Government Fails So Often." Whereas both studies claim that public policy failure is common (happening 'so often'), neither study bases its analysis on more than a limited number of specific cases, leaving one wondering 'how often' public policies really do fail?

Yes, failure happens, but is it 20% of the time, or 50%? And, what do we mean by failure, anyway?

These are important questions, because public policy results matter. Public policies are often needed to address society's toughest issues—where the market has failed, for instance, or where societies face collective action challenges, or where public goods need to be produced. The world is in trouble if public policies targeting such issues fail routinely, and we should know if this is indeed the case. Public policy initiatives also absorb a large portion of the world's resources—accounting for an estimated 16% of global gross domestic product in 2017,[1] or $13 trillion—and a high rate of policy failure would mean that we are wasting these resources. We need to know if this is the case, to consider reallocating our limited resources.

It is not easy to answer these questions, however. Governments—the main organizational entities pursuing public policies—seldom if ever provide data one can use to assess overall policy 'success' or 'failure' of chosen policy initiatives. One option they do offer is public sector deficit numbers, which allow one to see where governments spend more on policy engagements than citizens can afford. This measure does not capture what governments actually achieve with the money, however, through the many projects and activities they fund.[2] Beyond this, governments typically offer audit reports to report on spending behavior, but these tend to provide limited information on the compliance of spending behavior to legislated budgets.[3] Additionally, some governments produce performance data for some of their interventions, but the information is seldom comparable or accessible across the entire policy portfolio, to obtain a general picture of success or failure.[4] In the absence of such data, one could perhaps consider citizen surveys to see where government policies are perceived to succeed or fail.[5] Such surveys are hard to interpret,

[1] This statistic is obtained by calculating the average spending by governments across the world in 2017, as a percentage and in actual dollars. See https://www.theglobaleconomy.com/rankings/Government_size/

[2] This OECD description indicates how limited the deficit figure is in informing one of government performance (https://data.oecd.org/gga/general-government-deficit.htm).

[3] See, for instance, how spending requests for United States Federal agencies focus more on how agencies comply with legal budgets than how these agencies deliver effective policy products. https://www.gao.gov

[4] 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.

[5] The 2017 Edelman Trust Barometer found 43% of those surveyed trust Canada's government. Only 15% of those surveyed trusted government in South Africa, and trust levels were low in most countries—including Brazil (at 24%), South Korea (28%), the United Kingdom (36%), Australia, Japan, and Malaysia (37%), Germany (38%), Russia (45%), and the United States (47%). Similar surveys find trust in government averages only 40-45% across member countries of the Organization for Economic Cooperation and Development (OECD), and as few as 31% and 32% of Nigerians and Liberians trust government to do the right thing or get things done. See Edelman (2017). *2017 Edelman Trust Barometer*.https://www.slideshare.net/EdelmanInsights/2017-edelman-trust-barometer-global-results-71035413; See also, OECD (2013). *Government at a Glance, OECD Public Governance Reviews*, OECD Publishing, Paris: https://doi.org/10.1787/gov_glance-2013-en.; See, OECD (2017). *Trust and Public Policy: How Better*

however, and observers suggest that they reflect media-influenced frustration with economic and political conditions rather than experience or evidence-informed views on government policy success or failure.[6]

What we really need to assess 'how often' failure occurs is a database of policy results. Imagine if a public policy organization provided detailed, transparent, accessible data on the 'success' of all of its activities—projects, budget operations, and more—in a way that allowed deep-dive analysis of each intervention and broad reflection across the full portfolio of such interventions?

Luckily, such a database exists. The World Bank—a multilateral entity that pursues policy initiatives in and through over 150 governments across the world—assesses the success achieved in hundreds of unique projects every year. The Bank's internal ratings of such success are publicly available, providing a valuable lens into the performance of public policy interventions funded by the organization. This paper examines performance in over 400 of these interventions, all of which concluded operations in the period between 2016 and 2018—to determine the regularity of failure in the portfolio. The finding is that failure happens between 25% and 50% of the time—depending on how one defines success and failure.

The failure rate is about 25% if the definition focuses on what might be called 'project or product success'—where the policy intervention followed process norms and delivered on promised outputs (or near outcomes). The failure rate is about 50% if the definition asks whether the policy intervention solves the problem it was designed to solve, or is likely to produce more ambitious development outcomes or impacts (what I call 'problems solved, with development impact').

Through this analysis of failure rates, the study sheds light on what the Bank means by failure; is it just about project success or also about broader development impact? This allows a conversation of the topic for public policy organizations more generally, and especially for those organizations that develop and implement policy interventions using a 'plan and control' approach. The discussion suggests that these 'plan and control' organizations are likely biased to focus on 'project and product success' rather than 'problems are solved with development impact' because the former focal point is more amenable to 'planning and control' than the latter. This raises a new question in the conclusion, "Is this how citizens would want their public policy organizations to conceptualize success?" In suggesting a response to such question, the paper offers a new, blended approach to measuring success and failure.

Governance Can Help Rebuild Public Trust, OECD Public Governance Reviews, Paris: OECD. https://doi.org/10.1787/9789264268920-en; and Bratton, M. & Gyimah-Boadi, E. (2016). *Do trustworthy institutions matter for development? Corruption, trust, and government performance in Africa. Afrobarometer Dispatch No. 112.* http://afrobarometer.org/sites/default/files/publications/Dispatches/ab_r6_dispatchno112_trustworthy_institutions_and_development_in_africa.pdf.

[6] A critique of these polls is Nye, J., Zelikow, P. & King, D. (1997). *Why People Don't Trust Government*. Cambridge, M.A. : Harvard University Press. An example of such survey data is discussed in Clark, C. (2017). Americans Don't Trust Government and They're Angry About It, Poll Finds. *Government Executive*, December 18, 2017. https://www.govexec.com/management/2017/12/americans-dont-trust-government-and-theyre-angry-about-it-poll-finds/144646/

'Plan and Control' Public Policy Processes

At its simplest, public policy involves the many steps public organizations take to address problems raised by their constituents or members for attention. Policy interventions are made by an organization (or organizations) on behalf of the 'public' (or members), oriented toward a goal or desired state, such as the solution of a problem. These interventions are also typically part of an ongoing process without a clear beginning or end, since the challenges warranting policy attention—and the voices drawing attention to such challenges—are constantly changing.

Governments are the main organizations involved in making and delivering public policies. These governments serve their citizen publics. Other organizations also served constituency groups with public policy interventions, however. The World Bank is an example. It is an international multilateral organization that supports public policy interventions in over 150 member countries. Financing is the Bank's primary means of policy support, with the organization providing billions of dollars to hundreds of policy projects (also known as operations) every year. This paper examines the success and failure of the Bank in facilitating these policy engagements. Before looking at the Bank's success and failure record, however, it is important to reflect on how the organization works—and especially how it develops policy interventions.

The simple version of this story is that projects are identified, prepared and executed through a standard 'project cycle' used to structure policy engagements.[7] This cycle has multiple stages:[8]

- It begins with World Bank experts responding to a 'need' or 'problem' expressed by government officials in a particular country.
- These World Bank experts then engage expert officials in the relevant government if a process of identifying potential solutions to address such need.
- The World Bank staff (sometimes with government officials) then prepare a project proposal, deconstructing the solution into operational dimensions and a series of practical plans to guide implementation—including budget, procurement, and human resource plans, and a project timeline that specifies what will be done when, by whom, and at what cost..
- The project proposal then goes through appraisal, where a larger group of World Bank experts visit the country and do a due diligence analysis on the various plans.[9]

---

[7] See Awati, K. (undated). Are project management approaches generic? (https://www.projectsmart.co.uk/are-project-management-practices-generic.php) based on Besner, C., & Hobbs, B. (2008). Project management practice, generic or contextual: A reality check. *Project Management Journal* 39.1.: 16-33. See also Ika, L., Diallo, A, and Thullier, D. 2009. Project management in the international development industry: The project coordinator's perspective. *International Journal of Managing Projects in Business* Vol. 3 No. 1, 2010 pp. 61-93. See also Youker, R. (1988). *Managing the project cycle for time, cost and quality: lessons from World Bank experience*. https://www.pm4dev.com/resources/documents-and-articles/105-managing-the-project-cycle-robert-youker/file.html. See also Golini, R., Corti, B., & Landoni, P. (2017). More efficient project execution and evaluation with logical framework and project cycle management: evidence from international development projects. *Impact Assessment and Project Appraisal*, *35*(2), 128-138.

[8] World Bank. (1993). *The World Bank Project Cycle.* Washington, D.C.: World Bank. Available at http://documents.worldbank.org/curated/en/696601478501928227/The-World-Bank-project-cycle

[9] The World Bank teams ask various questions in this process, including the following (taken from the World Bank Information Brief on the World Bank Project Cycle). "Technical questions: Is the dam soundly designed and engineered? Does it meet acceptable standards? Will it displace local people? Will it affect the environment in any

- The project is then subjected to negotiations (between the World Bank and borrowing government) before being presented to and, potentially approved by, the World Bank board.
- Once approved, the project moves into implementation, where the host government and Bank officials identify an implementing agency to lead the work. A World Bank team provides supervision of the implementation process (regularly visiting to ensure the project is progressing as-planned) until the project is complete.
- An evaluation then ensues, where every project is subjected to a similar set of questions about results, process compliance, and risks to development outcomes.

This World Bank project cycle resembles project management processes employed by many governments and most public policy organizations working in the development field (including bilateral partners like the United States' Agency for International Development (USAID), the British Department for International Development (DFID), and others).[10] The project cycle is also similar to the budget planning and implementation processes commonly used to promote policy engagements in governments across the world.[11] These processes tend to begin with detailed up-front planning by line ministries (in devising policy submissions as budget requests). They move to a step where the plans are set into law (as budgets), and advance to a period of controlled execution where the goal is to ensure that implementation reflects what is in the budget plan. They finish with evaluations by external audit agencies that are sent to legislatures to ensure accountability for spending.

All of these mechanisms have a similar linear choreography (simplified in Figure 1) and are what I call 'plan and control' processes. They 'plan and control' processes start typically with a detailed planning stage (the World Bank's identification, preparation and appraisal steps), progress through a 'lock-in' moment where the plan is set in stone (the Bank's negotiation and Board approval steps), move to the implementation stage where various agencies are given responsibilities to execute, control, and supervise activities (with the goal of ensuring that agents

adverse way? Institutional questions: Does the borrower have the organization, the management, the staff, the policies to build a railroad and maintain it? If not, what changes are required to put these into place? Economic questions: Will the benefits of a water supply system outweigh the costs? What will its impact be on family incomes? Will it provide job opportunities for local people? What is the estimated rate of return on the investment? Financial questions: Is the borrower's financial plan sound? Is the electrical distribution system financially viable? Is the proposed accounting system adequate?" See this list of questions and further descriptions of the process at http://documents.worldbank.org/curated/en/696601478501928227/pdf/109412-BRI-WBG-PUBLIC-date-04-01-1993-The-World-Bank-Project-Cycle.pdf).

[10] Studies show that most policy organizations use these 'plan and control' mechanisms. See, for instance, Landoni, P. & Corti, B. (2011). The management of international development projects: moving toward a standard approach or differentiation? *Project Management Journal, 42*(3), 45–61. See, also, Muriithi, N., & Crawford, L. (2003). Approaches to project management in Africa: implications for international development projects. *International journal of project management* 21.5 (2003): 309-319.

[11] Studies show that countries are adopting these kinds of public financial management processes across the globe. See, for instance, Bryson, J.M., 2018. *Strategic planning for public and nonprofit organizations: A guide to strengthening and sustaining organizational achievement*. John Wiley & Sons. See also, Martí, C. (2018). Performance Budgeting and Medium-Term Expenditure Frameworks: A Comparison in OECD Central Governments. *Journal of Comparative Policy Analysis: Research and Practice*, pp.1-19.

follow the plan), and culminate in an evaluation stage (where questions are asked about how well the execution complied with plans and produced promised deliverables).

Figure 1. The linear progression of a plan and control public policy process

Plan			Lock-in	Implement	Evaluate
Identify	Prepare	Appraise	Negotiate, Approve	Execute, Monitor and Supervise, Control According to Plan	Evaluate Compliance to Plan

Rating Public Policy Success and Failure?

The World Bank has been 'rating' the projects it funds through these 'plan and control' processes for decades,[12] and publishing these ratings in reports that also go into detail on how projects were designed and implemented and how specific entities involved in the project performed (notably the World Bank teams and the borrowing government). The ratings are also compiled in a publicly available database, where one can view the basic assessment data associated with all past projects.[13] Links in the database allow interested parties also to access the detailed documents of each project, to pursue deeper analysis of unique project experiences.

The most referenced rating of each project—used to reflect might be called the project's 'success' in the organization's annual reports—captures how 'satisfactory' the project outcome was, from the Bank's point of view.[14] This rating captures the World Bank corporate perspective of a funded project's overall outcome, reflecting three main concerns. First, it assesses process compliance in executing a project (including if the project adhered to policy, budget and other requirements), or whether "there were [...] shortcomings in the operation's achievement of its objectives, in its efficiency or in its relevance" (World Bank, 2005). This is called 'efficiency'. Second, the rating speaks to "the extent to which the operation's major relevant objectives were achieved, or are expected to be achieved, efficiently" (World Bank, 2005). This is called 'efficacy'. Third, the rating

[12] Kapur, D., Lewis, J. P., & Webb, R. C. (2011). *The World Bank: its first half century* (Vol. 1). Washington, D.C.: Brookings Institution Press.

[13] A lot of research is based on these ratings, including the following: Bulman, D., Kolkma, W., & Kraay, A. (2017). Good countries or good projects? Comparing macro and micro correlates of World Bank and Asian Development Bank project performance. *The Review of International Organizations*, *12*(3), 335-363; Denizer, C., Kaufmann, D., & Kraay, A. (2013). Good countries or good projects? Macro and micro correlates of World Bank project performance. *Journal of Development Economics*, *105*, 288-302; Ika, L. A. (2015). Opening the black box of project management: Does World Bank project supervision influence project impact?. *International Journal of Project Management*, *33*(5), 1111-1123; Ika, L.A., Diallo, A. and Thuillier, D., 2012. Critical success factors for World Bank projects: An empirical investigation. *International journal of project management*, *30*(1), pp.105-116; Kilby, C. (2000). Supervision and performance: the case of World Bank projects. *Journal of Development Economics*, *62*(1), 233-259; Shin, W., Kim, Y., & Sohn, H. S. (2017). Do Different Implementing Partnerships Lead to Different Project Outcomes? Evidence from the World Bank Project-Level Evaluation Data. *World Development*, *95*, 268-284.

[14] The World Bank's corporate performance targets are based on this indicator, which are referred to as 'outcomes' in the World Bank's 2017 Results and Performance Report. (See World Bank. (2017). *Results and Performance of the World Bank Group 2017*. Washington, D.C.: World Bank (especially Chapter 2).

assesses whether a project's activities appear 'relevant' to the broad development goals or problems that led to the project need in the first place. This is called 'relevance'.

An ordinal six-point scale is used to capture this rating, showing how a project scores in the range from 'highly unsatisfactory' (a score of 1) to 'highly satisfactory' (a score of 6). Overall, the 'satisfaction' outcome rating will be high (a 5 or 6, denoting 'success') if a project produces its pre-defined objectives, and if these objectives were reached efficiently and appear relevant to achieve broader development outcomes. Bank reports tend to count any project that scores above 4 ('moderately satisfactory') as a 'success'.[15]

An example comes from the Georgia Regional Development Project, which was rated 'moderately satisfactory' given the following description:[16]

> "Relevance of the PDOs [Project Development Objectives] to the Government and Bank strategies is rated as High. Efficacy of the first objective - to improve the infrastructure services for supporting the development of a tourism-based economy - is rated Substantial. The efficacy of the second objective - to improve the institutional capacity to support the development of a tourism-based economy - is rated Modest, in terms of what was achieved. Overall efficacy, however, based on what was achieved together with what can be expected to be achieved in the next five to ten years, efficacy is rated substantial. Efficiency is rated Modest, based on the inefficiency of the project's cost effectiveness and given its administrative and operational inefficiencies, which contributed to delays during implementation and eventually non-completion of important activities at project closure."

This kind of World Bank project satisfaction rating does not directly assess whether policy interventions actually foster progress in addressing the broader development problems or needs that inspired project identification in the first place. Ratings merely suggest whether project deliverables are 'relevant' to these broader issues. The World Bank's Jurgen Blum[17] describes this limit to the World Bank's assessment protocol as follows:

> "For a given Bank project, IEG [Independent Evaluation Group] outcome ratings primarily measure "the extent to which the [project's] major relevant objectives were achieved" – not, or at least primarily not, whether the project made a difference for the client government's performance, or ultimately, for development outcomes. While "relevance" is one of the declared evaluation criteria, de facto evaluation practice suggests that the prime evaluation criterion for a project is whether it achieved its immediate objectives. Whether these are in turn relevant for achieving overarching development goals is often

[15] The only reference to performance indicators in the 2018 World Bank Annual Report, for instance, is to the percentage of projects considered 'moderately satisfactory' or better. A comment on 77 discusses the "*Results and Performance of the World Bank Group 2017* report " and notes that this report "found that 73 percent of World Bank projects completed between fiscal 2014–16 achieved moderately satisfactory or better development outcome ratings." (See World Bank. (2018). *Annual Report.* Washington, D.C.: World Bank).

[16] World Bank. (2018). *Georgia Regional Development Project. Project Implementation Completion Report Review*. Washington, D.C.: World Bank. http://documents.worldbank.org/curated/en/814031537536518091/pdf/Georgia-REG-DEV-1.pdf

[17] This section is taken from Blum's draft paper (2014, page 28). It was not included in the final World Bank working paper. 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.

contestable and may in practice play a much lesser role. It would be cynical to believe that many projects do not make a difference for broader development goals – but the extent to which this is the case and varies across projects is not reliably reflected in project ratings."

It is important to recognize that the Bank takes the challenge of producing broader development objectives through its projects seriously, however. The Independent Evaluation Group (IEG) tries to ensure that project preparation processes foster up-front thinking about 'relevance' by mandating task teams to present the 'theory of change' underpinning their operation in the preparation process.[18] The goal is to get teams to explain how they think the project will address 'the need' or 'problem' identified by government, and yield future outcomes that solve the problem.[19] Figure 2 illustrates a common approach to developing the theory of change in a project (based on what is commonly called the Logical Framework or the Results Framework).[20] According to this approach, project designers must identify the logical 'if-then-theory' behind a policy intervention.

Figure 2. The basic structure of a theory of change

A Logical or Results Framework/ Theory of Change Template	If input is provided	Then activity is undertaken	Then output is produced	Then outcome results	Then impact follows

Identifying the 'if-then-theory' starts with identifying the 'need' or 'problem' that should be addressed. For instance, a government might be concerned that 'firms are using outdated accounting practices that do not reveal the true value of debt, which puts our financial markets at risk'. The exercise then calls project teams to identify the inputs they will employ in a policy intervention. For instance, 'funds will be made available'. The 'theory of change' then moves to specifying activities that these inputs will produce, and explaining how and why the inputs are expected to yield such activities. For instance, 'the money in a project will pay for accountancy consultancies that will be procured through the government procurement process at specific dates and rates'. Then, the team has to identify outputs it expects to see after activities, and how and why these outputs are expected. For instance, 'the consultants

[18] See World Bank. (2012). *Designing a Results Framework for Achieving Results.* Washington, D.C.: World Bank. (https://siteresources.worldbank.org/EXTEVACAPDEV/Resources/designing_results_framework.pdf). Also Bisis, P. *Theory-of-change-vs logical-framework?* (http://www.tools4dev.org/resources/theory-of-change-vs-logical-framework-whats-the-difference-in-practice/). And Britain's Department for International Development (DFID): https://assets.publishing.service.gov.uk/media/57a08a66ed915d622c000703/Appendix_3_ToC_Examples.pdf

[19] Interestingly, a 2017 evaluation found that less than a third of World Bank projects include an explicit theory of change. See the Brettonwoods Project blog post, *IEG critical of World Bank's progress on shared prosperity goal* (https://www.brettonwoodsproject.org/2018/03/ieg-critical-banks-progress-shared-prosperity-goal/)

[20] For background on the logical framework approach, see Bell, S. (2000). Logical Frameworks, Aristotle and soft systems: a note on the origins, values and uses of logical frameworks, in reply to Gasper. *Public Administration and Development* 20(1):29 – 31. See also Golini, R., Corti, B., & Landoni, P. (2017). More efficient project execution and evaluation with logical framework and project cycle management: evidence from international development projects. *Impact Assessment and Project Appraisal*, *35*(2), 128-138. Some question if the logical framework is a 'theory of change' method. (https://www.annmurraybrown.com/single-post/2016/03/20/Theory-of-Change-vsThe-Logic-Model-Never-Be-Confused-Again). The logical/results framework is used widely, however. See povertyactionlab(https://www.povertyactionlab.org/sites/default/files/3.%20Theory%20of%20Change%202014.03.10.pdf). See also World Bank. (2012). *Designing a Results Framework for Achieving Results.* Washington, D.C.: World Bank. (https://siteresources.worldbank.org/EXTEVACAPDEV/Resources/designing_results_framework.pdf).

will produce a new draft accounting law that reflects international best practice'. The theory of change then specifies outcomes expected from outputs, and explains how and why those outcomes are expected. For instance, the draft law will become a formal law, which will mean that government has new accounting legislation in place to mandate the adoption of improved accounting practices in firms. Finally, the theory of change explains how proposed outcomes are expected to yield needed impacts—most clearly, solving the problem or addressing the need expressed by government. For instance, 'the new law will force firms to adopt improved accounting methods and this will improve the way they account for debt, which will decrease the risk of unaccounted debt in our market'.

The theory of change approach is intended to help project designers determine the scope of operational commitment, structure the operation, and communicate how the operation is (theoretically) expected to yield broader development impacts (and solve whatever problem the government is concerned about).[21]

Given such reasoning, evaluators often concern themselves with how valid the theory of change is, and whether there is a strong or weak likelihood of jumps from the activities and outputs produced directly in a project and the outcomes and impact needed by countries.

This concern is reflected in the Implementation Completion Review (ICR) Report for the Bolivian Decentralized Electricity for Universal Access project, where the independent evaluators noted, "While the link between the project's inputs and outputs is clear, project outcomes were not defined and therefore the link between the actual outputs and outcomes and the PDO is missing."[22] This evaluation concern is captured explicitly in an additional rating used by the World Bank—the 'Risk to Development Outcomes' (RDO). This is described as "the risk, at the time of evaluation, that development outcomes (or expected outcomes) will not

[21] A Theory of Change in a World Bank project is Nigeria's Public Sector Governance Reform and Development Project (World Bank. (2017). *Nigeria Public Sector Governance Reform and Development Project Implementation Completion Report Review*. Washington, D.C.: World Bank. http://projects.worldbank.org/P097026/nigeria-state-governance-capacity-building-project-ii?lang=en&tab=documents&subTab=projectDocuments). This Theory of Change reads as follows: "The modernization of the public financial management systems (PFMs) was to be achieved through project-supported capacity building activities focused on accounting and financial reporting, preparing budgets, conducting internal and external audits, improving tax administration, and enhancing monitoring and evaluation (M&E) systems, as well as through the implementation of a modern legal and regulatory framework for public procurement. Capacity building activities for the personnel and payroll systems staff of the respective Ministries, Departments and Agencies (MDAs) and the modernization of personnel records in the state public service were expected to strengthen human resource management. It was also anticipated that, taken together, these activities would enhance transparency in the use of public resources in the Tier I states. Installing the Budget and Treasury Management System (BATMIS) and technical assistance support for budget preparation, procurement, human resource management and strengthening M&E, was expected to improve transparency in public finance and human resource management in the states which had implemented the previous First State Governance and Capacity Building Project. Greater transparency in public financial and human resource management in the Tier II states was expected to be achieved by targeted technical assistance activities."

[22] World Bank. (2014). *Bolivian Decentralized Electricity for Universal Access Project Implementation Completion Report Review*. Washington, D.C.: World Bank.

be maintained (or realized)."[23] Each project is rated for RDO on an ordinal four-point scale that ranges from 'low' (a score of 1) to 'high' (a score of 4). In the Bolivian project, the RDO was 'significant' (a 3) given concerns that project outputs would not yield more results because of "economic and social instability, political interference, low demand [or the products delivered through the project] and lack of interest from the private sector" as well as concern over who would continue implementation once the project was completed.[24]

This Risk to Development Outcomes rating is an additional piece of information to consider—with the project satisfaction rating—when evaluating World Bank project success or failure.

Assessments of these risks and satisfaction outcomes emerge from a multi-stage process.[25] They are conducted initially by the Bank's internal Task Teams (who managed the projects in the last throes of implementation). Ratings from these teams are captured in documents called Implementation Completion Reviews (ICRs). The World Bank's Independent Evaluation Group (IEG) reviews these assessments in desk-based studies to offer a second-hand validation (where the ICR ratings are examined in light of a review of past project documents and selected interviews with Bank team members). The IEG also selects a sample of the projects to review in more detail each year, visiting the countries to interview government officials, review results on the ground, and provide a first-hand perspective on performance.

There are many critiques one can make of these assessments; they do not represent a broad view on results, for instance, and are certainly not of academic quality in determining validity of reliability or even comparability in findings. In fairness, the Bank has never claimed any of these qualities in reporting on the ratings. They use the ratings to communicate a corporate perspective on performance across their portfolio of activities, to showcase overall results to shareholders and to inform internal thinking about where the Bank works, how it works, and with what impact. This study uses the ratings in a similar way, not as objective measures of success or failure but rather as signals provided by the Bank of its own views on its own success and failure.

'How Often' Do Public Policies Fail? Well, It Depends…

This study looks at the assessments of 416 World Bank projects rated in the period between 2016 and 2018, to first get a glance of the Bank's own 'performance picture' and in particular how often the Bank—as a major public policy organization—thinks about failure. The sample is limited to years post 2015 to capture changes in coding introduced in that year. This sample is still impressive, however, covering projects in over 80 countries and activities in many sectors common to public policy interventions—as shown in Table 1.

[23] 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

[24] Ibid.

[25] Blum provides a good description of the assessment process. 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.

The projects were all completed in the years between 2016 and 2018, and accounted for $88 billion in total, with allocations ranging between $2 million and $2 billion. The projects were between 1 and 12 years in duration (the longest project began in 2004 and finished in 2016).

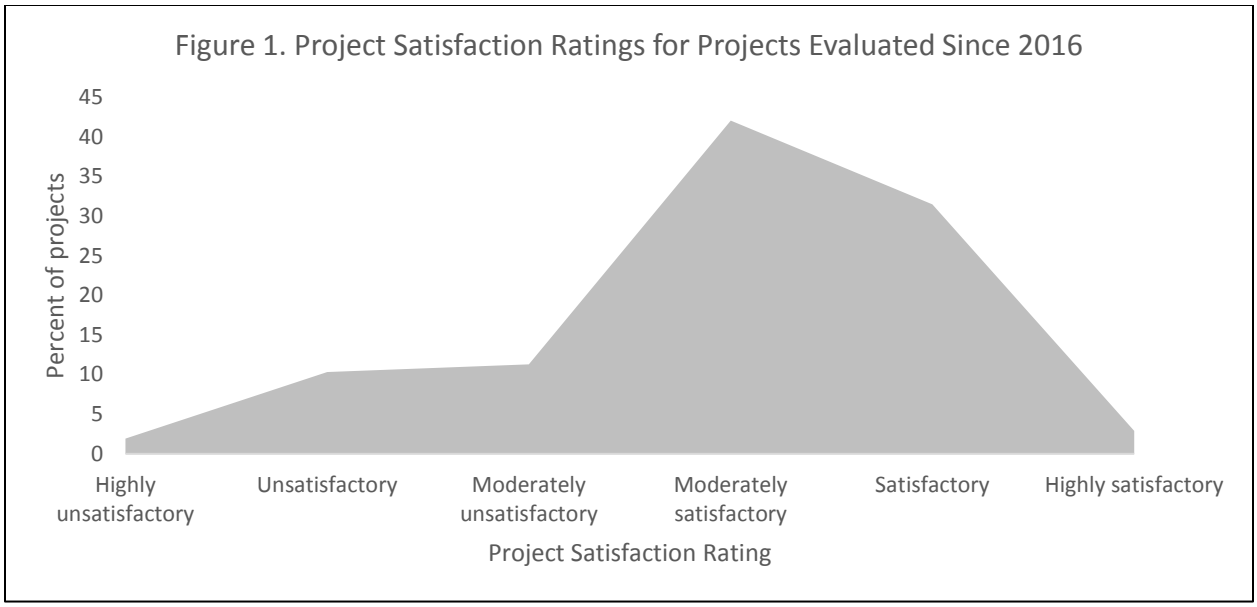
Table 1. Sectors represented in the sample of World Bank projects

Sector	**Number of Projects**
Agriculture and rural development	64
Transport	47
Urban	37
Energy and mining	33
Environment	32
Education	31
Health, nutrition and population	25
Water	24
Financial and Private Sector Development	21
Public sector governance	20
Financial and Private Sector Development	20
Social protection	16
Economic policy	15
Global Information/Communications Technology	6
Competitive Industries Practice	2
Financial Management	2
Investment Climate Practice	1

The variation of projects in this database suggests that one could indeed be looking at the public policy portfolio of a government. So, what does failure look like in that portfolio?

Looking at 'Project Satisfaction' Results

As explained, the World Bank 'project satisfaction' rating system allows for more than a 'thumbs up or thumbs down' view on the success or failure of each project. Data show levels of satisfaction from 1 (highly unsatisfactory) to 6 (highly satisfactory), which means that we can get a view on the regularity with which Bank projects achieve different levels of success (or failure). As shown in Figure 1, this view suggests first that extreme public policy failure is extremely rare (fewer than 2% of the projects received 'highly unsatisfactory' ratings). More general public policy failure (where a project receives an 'unsatisfactory' rating) is more common—happening 10.3% of the time—and more qualified failure (where a project receives a 'moderately unsatisfactory' rating) is even a bit more common—happening 11% of the time. Altogether, projects falling into these lower categories—of qualified to extreme failure—account for 24% of all projects.



Source: World Bank projects database.

After accounting for such 'failures' one is left with 76% of the projects—over three quarters of all the public policy initiatives funded by the Bank in the period in question—in more positive assessment categories. Only 3% of these projects were considered 'highly satisfactory'—akin to being extreme successes—but 31% were assessed as 'satisfactory'—akin to being a full success. A further 42% were assessed as 'moderately satisfactory'—similar, perhaps, to being called a qualified success.

These results suggest the following on the regularity of failure and success in public policy initiatives, at least in initiatives supported by the World Bank:

- Projects fail outright (or worse) 10.5% of the time.
- Projects are fully successful (or better) 34% of the time.
- Projects achieve mixed results 53% of the time, and in most of those instances (about 80% of the cases) the results tend to reflect more success than failure.

Whilst this view of World Bank project performance is subject to the caveats raised already about the quality of data, and requires much more interpretation and analysis, it surely indicates an answer to the question asked at the start of this paper: public policy failure is not that common and does not happen 'so often' as some suggest. Indeed, the evidence indicates that public policy success is more common than failure, but the most common outcome is qualified progress.

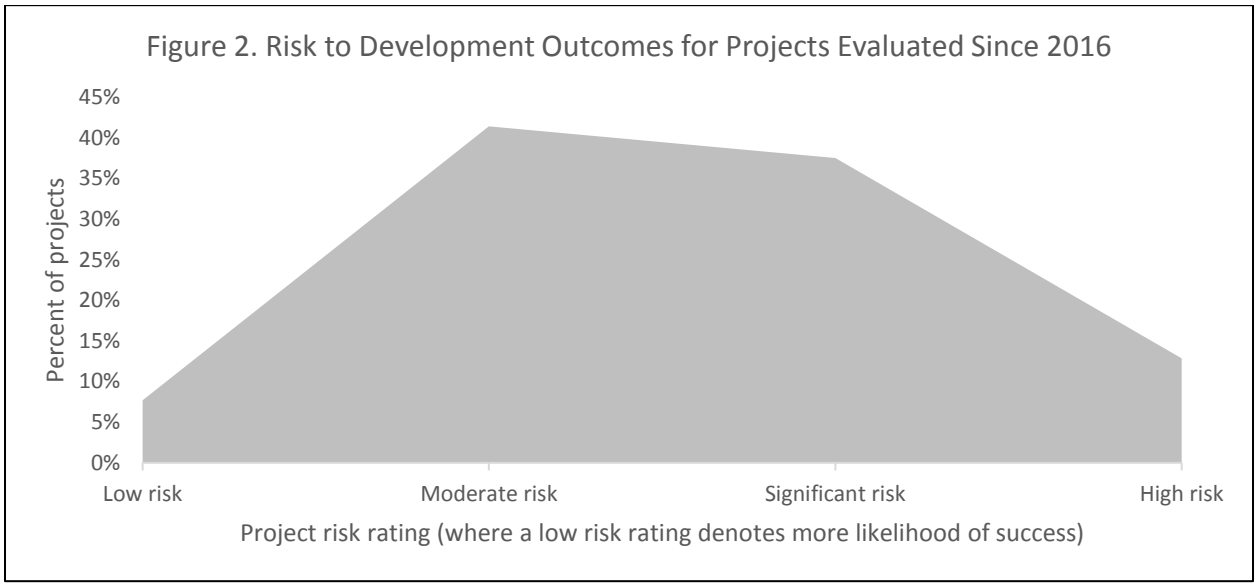
Looking at 'Risk to Development Outcome' Results

This statistic is not the end of the analysis, however, even with regards to the basic information provided in the World Bank project database. As noted, the 'project satisfaction rating' only assesses what the Bank project intended to deliver in the initial project plan, and the efficiency of delivery (or extent of control in the execution process). It does not directly assess the development impact of the public policy intervention made through the project—whether the

project's deliverables solve whatever problem originally motivated the project, or if the project's deliverables were used, or fostered real development. In Blum's words, these ratings do not capture "whether the project made a difference for the client government's performance, or ultimately, for development outcomes."[26]

As discussed, the World Bank does try to capture some view of this kind of impact, however, through its rating on the 'Risk to Development Outcomes' (RDO). As noted, this is described as "the risk, at the time of evaluation, that development outcomes (or expected outcomes) will not be maintained (or realized)." As also noted, each project is rated for RDO on an ordinal four-point scale that ranges from 'low' (a score of 1) to 'high' (a score of 4). A more successful project would score a low rating—1—because evaluators think that there is a low risk to the achievement or maintenance of outcomes. A less successful project (or one that might be called a 'failure risk') would score a high rating—3 or 4—because evaluators think there is a high risk to achieving or maintaining outcomes.

Figure 2 sums the ratings on RDO for 387 of the 416 projects in the database, which tell a different story to the 'satisfaction' ratings.[27] Projects with 'high risk' are at the extreme right, and account for 13% of the total. These are the projects with greatest chance of failure—where evaluators worry about the lack of or non-sustainability of development outcomes. Projects with 'significant risk' are next-to-the-right-extreme and account for 38% of all projects. Together, this is the significant to high risk of failure project group. It accounts for 51% of the public policy portfolio.



This leaves 49% of the projects in the 'low risk' to 'moderate (or mild) risk' categories. These are the projects one could categorize as 'more likely to succeed' (low risk) or 'fairly likely to succeed, with qualification'. The more positive of these categories only includes 8% of the projects in the World Bank's portfolio. The second, more qualified category, includes 41% of the projects.

[26] 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. (his draft paper, page 28).
[27] Note that some projects were missing data on RDO.

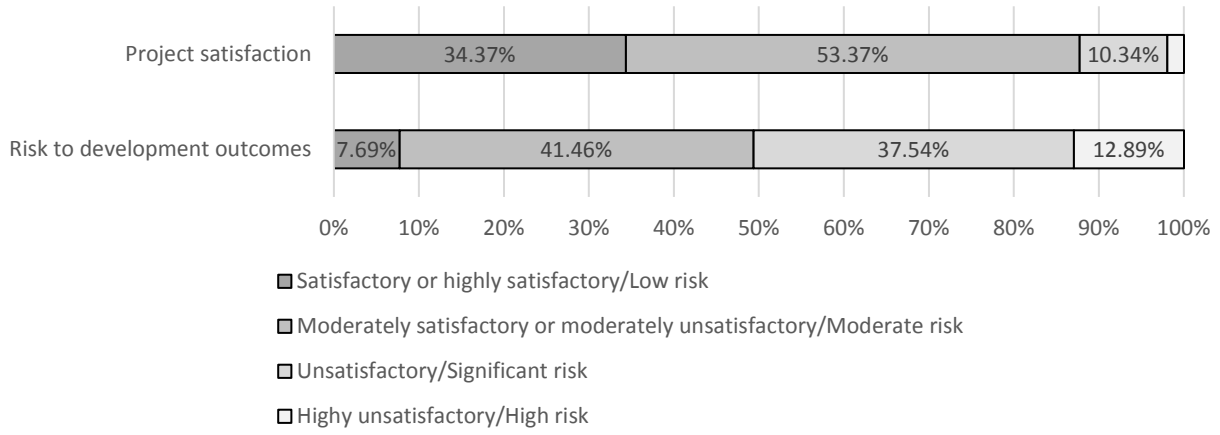
The headline of a story about these ratings could quite easily read something like '51% of World Bank projects are at significant to high risk of failing to foster development outcomes'. This is quite a different message of public policy failure to that derived from 'satisfaction' ratings, where projects in the three most pessimistic categories (highly unsatisfactory, unsatisfactory, and moderately unsatisfactory) only account for 24% of the total portfolio.

The Differences in Failure Rates

An emerging part of this narrative on policy failure centers on the differences between the two 'project results' measures captured by the World Bank—project satisfaction and risk to development objectives. Figure 3 summarizes the differences in performance reflected in these two ratings. The figure was developed after transforming the project satisfaction data from a six point scale to a four point scale, allowing direct comparison with the RDO data:

- To do this, the 'highly satisfactory and 'satisfactory' categories were merged, to compare with the 'low risk' RDO category (as categories capturing projects with more successful attributes).
- The 'moderately satisfactory' and 'moderately unsatisfactory' categories were also merged, to compare with the 'moderate/modest risk' RDO category (as the two categories capturing projects considered 'qualified' or neither successes nor failures).
- The final two categories are more negative, in the 'failure' territory—'unsatisfactory' and 'significant risk' and 'highly unsatisfactory' and 'high risk'.

Figure 3. Contrasting Project Satisfaction and Risk to Development Objectives Ratings for Projects Evaluated Since 2016



One could consider other ways of re-organizing and comparing these two rating sets, but the comparison would probably look quite similar. This comparison shows that satisfaction results are significantly better than RDO results. The proportion of successes is higher in respect of the satisfaction ratings, with 34% of projects falling in the 'satisfactory and highly satisfactory category' and a further 53% considered 'moderately satisfactory or moderately unsatisfactory'. In contrast, only 8% of projects were considered low risk and only 41% were considered moderate or modest risk. On the other hand, the proportion of failures (or potential failures) is

higher in respect of the risk to outcomes ratings—13% of projects were 'high risk' as compared with 2% of 'highly unsatisfactory projects'; 38% of projects were 'significant risk' as compared with 10% of projects considered 'unsatisfactory'.

The figure illustrates how contingent the rate of World Bank public policy failure is on the rating one looks at. This failure rate depends on whether one is considering direct deliverables of a current intervention (project satisfaction) or the potential of those deliverables have to yield indirect future outcomes (the risk to development objectives).

Determining Policy Failure Rates: A 'Plan and Control' Bias?

So far, this analysis shows—at least in the context of the World Bank—that public policies do not fail 60% or 80% of the time. The failure rate is, rather, between 20% and 50%—depending on which of the World Bank's assessment measures one focuses on. Which raises the question: which of the two measures should one focus on—satisfaction with immediate project performance or risk to producing or maintaining more demanding outcomes?

The answer inside the World Bank, is clear: satisfaction rates matter—not the risk to development outcomes. This is demonstrated in how the Bank reflects on its success. Consider a 2018 IEG comment on the question; 'Are World Bank Group Projects Getting Results?'[28]

> "Each year, IEG reviews the aggregate development effectiveness of World Bank Group projects completed during the past three years. Have projects delivered measurable results for their clients – and are the World Bank Group's constituent institutions meeting their corporate targets? The 2017 Results and Performance of the World Bank Group report answers some of these questions for projects completed during the FY14-16 period. We found that, although the share of World Bank (IBRD & IDA) projects rated moderately satisfactory or above (MS+) increased by three percent since the FY11-13 period, it remained below the corporate scorecard target of 75 percent."

A key take-away from this is that the corporate scorecard target for World Bank success centers on the proportion of projects considered 'moderately satisfactory' or above (MS+). This is obviously self-serving, with the organization focusing on the lower of two failure rates; it is easier to say 'we fail 24% of the time' than it is to say 'we fail 51% of the time.' However, the focus on 'project satisfaction' results is not just about the World Bank's incentive to tell the best possible performance story. This kind of focus is common for policy interventions introduced through the 'plan and control' approach described earlier in this paper. Interventions produced through this dominant approach tend to be biased towards looking at the immediate project product and how this product came about, not impacts of the product.

The bias is perhaps best understood when considering how 'plan and control' interventions operate in context of the 'theory of change' discussed earlier. As introduced, theories of

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

change used in project preparation and in budget proposals—especially those in the logical and results framework tradition—commonly specify the links between policy inputs and activities, activities and outputs, outputs and outcomes, and outcomes and impacts. The outcomes and impacts are arguably what governments and citizens want from policy interventions, but most projects execute this theory of change up to the output or initial outcomes stages only. The project deliverable is, therefore, defined at this stage—and the project commits to produce this deliverable.[29]

Using an example from earlier in this paper (about a project addressing debt issues in the private sector), the focus on products and processes could involve a commitment to deliver a draft new accounting law (with financial inputs procuring consultant activities and these leading to a key output—the draft law). The project would not commit to further results, which would require engagements and actions that are not under the direct control of the project designers or implementers (legislators turning the draft law into formal legislation, for instance, and private firms complying with this law). These outcomes and impacts are achieved in the 'aftermath' of the project, and the likelihood of achieving such depends less on the direct work of project and more on the 'theory of change' on which the project is based.

An example (from the World Bank projects) of this bias to evaluate inputs and activities and outputs (with light attention to outcomes) and not to focus on outcomes and impact is the 2012 Balochistan Disaster Management Project evaluation. The evaluation was undertaken with a limited focus on assessing performance against the "Project Development Objective as formulated in the Grant Agreement." This objective was "to strengthen the capacity of the Project Implementing Entity to prepare for and respond to natural disasters (Grant Agreement, page 5, Schedule 1)."[30] Within this focus, the project was considered satisfactory (scoring 5 out of 6 on the World Bank's success system) given that "17 of the output targets were achieved or exceeded,[31] only 3 were not met and one was awaiting approval at project closure." In addition, three 'outcomes' were also considered as met ('Increased awareness of the hazards and risk environment in Quetta', and 'Improved disaster response mechanism at PDMA Balochistan' and 'Enhanced capacity at PDMA Balochistan for implementing Community Based Disaster Risk Management Initiatives').[32]

However, the project was considered 'high risk' when it came to reaching development outcomes, because "the success of the efforts to sustain the achievements of the project will

[29] See Baccarini, D. (1999). The Logical Framework Method for Defining Project Success. *Project Management Journal* Volume 30(Issue 4):25-32.

[30] World Bank. (2016). Balochistan DM Project Implementation Completion Report Review. Washington, D.C.: World Bank, pg 1. http://documents.worldbank.org/curated/en/460941505920549900/pdf/ICRR-Disclosable-P127253-09-20-2017-1505920540311.pdf

[31] These outputs included "15 master trainers trained; 4 training sessions for responders in selected union councils in Quetta; 10,000 instruction materials on CBDRM available for dissemination to communities; 5 training sessions conducted by master trainers in selected union councils in Quetta; Standard Operating Procedures for PDMA on disaster response prepared; Emergency communications equipment operationalized; Balochistan DRM plan 2008 reviewed and updated; Capacity assessment undertaken and capacity enhancement plan prepared; Logistics plan developed for emergency stockpiles in selected districts; 2 pilot stockpiles established in Quetta based on logistics plan; Provincial CBDRM Program Developed.

[32] Ibid, pages 4-5.

depend upon the continued commitment of the Government of Balochistan to the Disaster Risk Management (DRM) agenda, which remains uncertain." A list of specific risks to reaching or maintaining development outcomes includes, "[The] risk that the capacity enhancement plan will not be implemented";[33] "[The] risk that the Disaster Management Information System (DMIS) will not be maintained";[34] "[The risk] that government's commitment and sense of ownership [we]re dependent on individuals and there is a high risk that this may change with administrative and political changes in government"; "[The risk that] further DRM planning and implementation may be challenged by the volatile security situation."

In sum, the project is included in the 'success' part of the World Bank's portfolio because it delivered the outputs promised in the grant agreement; even though the evaluators seem extremely pessimistic that the products delivered will actually be implemented or yield outcomes or impacts needed. 'Success' in this approach is about delivering what one can plan and control, even if this does not assure achievement of real development outcomes (which are relegated to the project's aftermath, and the vagaries of politics and policy uncertainty).

This 'plan and control' limit to thinking about project success is shown in Figure 4.

Figure 4. Plan and Control processes, World Bank ratings, Theory of Change, and Evaluation Limits

<table>
<tr><td>The Plan and Control Project process</td><td>Plan</td><td>Lock-in</td><td colspan="2">Implement</td><td>Evaluate</td><td>Aftermath</td></tr>
<tr><td>A Theory of Change Template</td><td colspan="2">If input is provided</td><td>Then activity is undertaken</td><td>Then output is produced</td><td colspan="2">Then outcome results</td><td>Then impact follows</td></tr>
<tr><td rowspan="2">World Bank ratings concerns</td><td colspan="3">1. Project Process Compliance (efficiency)</td><td colspan="2">2. Project objectives achieved</td><td colspan="2" rowspan="2">3. Relevance for broader development outcomes (Assessed in 'Risk to Development Outcomes' ratings)</td></tr>
<tr><td colspan="5">Both assessed in the 'Project Satisfaction' ratings</td></tr>
<tr><td>Common project evaluation dimensions</td><td colspan="3">Process and Project Management Success</td><td colspan="2">Product Success</td><td colspan="2">Social and Organizational Success
(Business and Strategic Success)</td></tr>
</table>

[33] The report notes that, "It remains uncertain whether the government will create or regularize the proposed key positions in the capacity enhancement plan. The decision to revise provincial civil service structures is complex and requires high-level political commitment which may not be forthcoming. PDMA [the implementing agency] is hiring staff on contract basis and has purchased equipment as detailed in the capacity enhancement plan, but no mechanism has been developed for its maintenance."

[34] The report notes that, "The PDMA [the implementing agency] has not institutionalized a system for regular hazard risk assessment that can feed into the DMIS [the new implementation system]. The project conducted one hazard and risk assessment for Quetta district only, but there is a need to conduct similar assessments for other disaster-prone areas of Balochistan to ensure comprehensive coverage of the province. There is also no system to track population changes and increases and decreases in public infrastructure, which are essential for improved planning."

The top row of this figure should be clear, as it is a reproduction of the top row in Figure 1, and simply shows the various elements in a 'Plan and Control' project process. The second row shows the 'logical if-then' parts of a standard logical or results framework theory of change. The key message in considering both rows is that projects typically act on the theory of change until the point of 'outputs' (and sometimes initial outcomes), which leaves 'if-then' assumptions and arguments about moving from outputs to outcomes and impacts untested—and delegates the execution on these links to actors and periods after the project is completed.

The third row of Figure 4 shows how the World Bank assesses its performance on individual projects, in the two key metrics discussed—'satisfaction' and 'risk to development outcomes'. The key message here is that the 'satisfaction' rating measures compliance and delivery that occurs within the project itself, and in the input to activity to output (and, perhaps, initial or proximate outcomes[35]) part of the 'if-then' theory of change. This is the part of the theory of change directly influenced by the project intervention (that can be planned for, and controlled).

The 'risk to development outcome' rating speaks to concerns about results that occur (or do not) in the aftermath of the project—where the theory of change considers more advanced outcomes and impact. The project cannot be held directly accountable for these kinds of results, because of a number of gaps and external factors: a time gap is crucial here, and manifests in the period between when the project ends and when advanced outcomes and impacts occur; management and control gaps also matter, given that new entities take over these roles once the project is completed; leadership gaps pertain to the change in political oversight that happens once projects end; external factors that cannot be planned for or also controlled typically fester between the direct production of outputs in a project and the experience of indirect outcomes and impacts, including political and economic change, organizational adjustment, and more.

To further illustrate this point, the bottom row in Figure 4 reflects theoretical work that has been done to explain and describe the various dimensions of project 'success'. There is a large literature on this topic, which argues consistently that there are multiple ways of measuring project success. A common approach delineates between 'Process and Project Management' Success, 'Product' Success, and 'Social and Organizational (or Business) Strategic' Success:[36]

- Process and Project Management Success measures the immediate performance of a project against its main design parameters—schedule (time), budget (cost), scope, and/or quality (the iron triangle, or three-legged stool of project management). Bannerman notes that this concern "remains the most widely used measure of project success" and notes that, "[I]ts

[35] The idea or proximate outcomes reflects the view that outcomes are often part of the project deliverables only when there is a close logistical connection between outputs and the next step outcomes. For instance, if the output is an information system and the outcome is 'data entered into the system' or where the output is 'roads built' and the outcome is 'deliveries made on new roads.' These are the 'proximate' outcomes, but they are not the more advanced outcomes one might be considering; like 'IT systems are actively used to make decisions' or 'deliveries made on new roads are more cost efficient and yield higher levels of trade'.

[36] See, for instance, Bannerman, P. L. (2008). Defining project success: A multilevel framework. In *Proceedings of the Project Management Institute Research Conference* (pp. 1-14). Also, Gemünden, H. G. (2015). Success factors of global new product development programs, the definition of project success, knowledge sharing, and special issues of project management journal. *Project management journal*, *46*(1), 2-11.

main value is in offering a simple, direct measure of performance of a project and the project management expertise applied to complete the project within the bounds of the most immediate design parameters (time, cost, and scope)."

- Product success measures the extent to which a project delivered its promised 'products', and if those products were used and considered useful by intended users (or provided intended benefits to targeted beneficiaries). As with process and project management success, this measure can be assessed directly against the design parameters in the project plan—where evaluators can ask whether stated 'deliverables' were actually delivered (according to planned specification, quality considerations, and the like).
- Social and Organizational (Business) and Strategic Success assessments go beyond the design parameters of the project plan, to ask whether interventions solved the particular problem (albeit in a timely, cost-efficient, and effective manner) that warranted an intervention in the first place, and—even more expansively—if the project better positions the community or organization to address future problems or take future opportunities and benefits (by ensuring, for instance, that capabilities are improved through the project).

Figure 4 suggests a direct overlap between the 'process and project management' and 'product success' concepts and the World Bank's 'satisfaction' ratings—and with theory of change deliverables up to outputs and proximate outcomes. These are the key 'success' measures for any 'plan and control' policy intervention—because they can be planned for and controlled. Concerns about 'social and organizational (business) and strategic' success are seen to overlap with the more demanding development outcomes concerns in the Bank. As discussed, these are captured by the 'Risk to Development Outcomes' rating, and might be called 'problems are solved with development impact' success in the public policy context. This is akin to achieving outcomes and impacts in a common theory of change, which usually unfolds in the aftermath of a typical World Bank project (and many other public policy interventions). Plan and control interventions (like World Bank projects) seldom focus on these results because they occur indirectly and after the project is done, are subject to external factors and gaps, and are hard to 'plan and control'.

Policy Failure: Often Enough, and Defined Too Narrowly

This paper asks a basic question: 'How often do public policy interventions fail'? It does so by examining over 400 projects in the World Bank's policy portfolio. The Bank is an organization that supports policy interventions found in governments across the world and that provides data on the success and failure of these interventions.

The study finds that there are different answers to this 'how often' question, depending on responses to a second question: 'What is failure anyway?' The failure rate is about 24% when one asks about 'project and product success' (whether planned products are delivered through an efficient process). The failure rate is 51% when one takes a broader view of results and asks if 'problems are solved with development impact.' ("Do policies solve the problems that warranted intervention in the first place, or whether the policies promoted development outcomes?")

This evidence suggests that failure is at either a manageable level (24%) or a high level (51%) depending on how one conceived of policy failure. It notes that the World Bank uses the lower

figure when it reflects on its own failure (and publishes such to the world at large). The study argues that this tendency to focus on 'project and product success' is not just a World Bank habit. It is likely the approach taken in any public policy organization that uses 'plan and control' management methods. These organizations will define success in terms of whether their 'plans' were well implemented, given strong 'control'. Any results that cannot be planned for, or controlled, are considered outside of the influence of the policy organization.

So, one can expect organizations like the World Bank and governments across the globe to declare success if their policy interventions produce new roads, laws, organizations, even research papers and training programs—all promised in plans and budgets and project agreements—even if no one drives on the roads, or complies with the laws, or works with the organizations, or reads the research papers, or learns anything new in the training programs.

Is this how citizens would want their public policy organizations to conceptualize success?

The chances are that it is not. Citizens and other constituencies—business communities, or even governments in the case of organizations like the World Bank—engage their public policy organizations with real problems that require real solutions. These solutions are only really manifest when real outcomes and impact are achieved (the success criteria I call, 'problems are solved with development impact'). The fact that these results are often achieved after projects are finished does not mean that projects should be 'off the hook' in ensuring they are achieved—or at least in managing risks to their achievement such that the 'risk to development objectives' is low at the point of project completion.

At present, the World Bank's project portfolio is dominated by policy interventions that end up with 'significant to high' risks to development objectives (51% of the projects are in this category). These 'high risk' projects should surely not be seen as successes, even if they met limited project delivery objectives (and are considered 'project and product successes')?

Figure 5 provides a broader way of thinking about success and failure in the World Bank and other 'plan and control' organizations, that is sensitive to the constraint of time-sensitive project delivery and to the need for policy interventions to actually have outcomes beyond such projects.

Figure 5. Merging different perspectives on policy failure, for a broader picture

	Highly unsatisfactory, unsatisfactory, or moderately unsatisfactory results	Moderately satisfactory, satisfactory, or highly satisfactory results
Low to moderate risk to development outcomes	Q2 (5%)	Q4 (42%)
Significant to high risk to development outcomes	Q1 (22%)	Q3 (31%)

Source: World Bank project database. Note that the numbers here do not sum to the same as numbers in Figures 1 and 2 as only 387 projects are included here, given missing data for some of the 416 projects.

Figure 5 merges both ratings to show the proportion of projects in four different quadrants:

- The first quadrant (Q1) is the most negative. It includes projects that were considered general to extreme failures in the 'satisfaction' rating (moderately unsatisfactory, unsatisfactory, or highly unsatisfactory) and were also seen as very risky in terms of the potential to deliver development outcomes (with significant or high risk to development outcomes). These projects—22% of the portfolio—were considered failures on both metrics. They should most worry a public policy organization like the World Bank, which cannot seem to deliver outputs or outcomes or impacts in the relevant project areas.
- The second quadrant (Q2) is also concerning. It includes projects that were considered general to extreme failures in the 'satisfaction' rating (moderately unsatisfactory, unsatisfactory, or highly unsatisfactory) but were not considered risky in terms of the potential to deliver development outcomes (with modest or low risk to development outcomes). These projects—5% of the portfolio—are curious to consider, as the public policy organization could not deliver outputs satisfactorily but is not concerned about emerging outcomes and impacts.
- The third quadrant (Q3) is also concerning. It includes projects that were considered general to extreme successes in the 'satisfaction' rating (moderately satisfactory, satisfactory, or highly satisfactory) but risky in terms of the potential to deliver development outcomes (with significant or high risk to development outcomes). These projects—31% of the portfolio—are those in which the public policy organization has been able to deliver immediate outputs but is very uncertain about delivering actual solutions to policy problems, or outcomes and impacts that improve development.
- The fourth quadrant (Q4) is the most positive. It includes projects that were considered general to extreme successes in the 'satisfaction' rating (moderately unsatisfactory, unsatisfactory, or highly unsatisfactory) and were also seen as not risky in terms of the potential to deliver development outcomes (with low or modest risks to development outcomes). These projects—42% of the portfolio—were considered successes on both metrics. They should most please a public policy organization like the World Bank, which seems able deliver outputs, outcomes and impacts in the relevant project areas.

Figure 5 suggests a broader view on public policy success and failure that currently exists in the World Bank—or, I believe, in any organization working in a plan and control tradition that tends to foster a narrow 'project and product success' view. This new view suggests that only 41% of the World Bank's policy interventions are both direct project and product successes and have the potential to foster indirect development outcomes and impacts. A larger portion of the organization's policy portfolio—59%—is failing on either the direct 'project and product' performance measure or the indirect 'problems are solved with development impact' performance measure, or both.

This is a public policy failure rate that is surely 'too often'; we need policies that more regularly solve problems, and that use resources more effectively in so doing. This raises a parting question: "What can be done to decrease the failure rate, and push more public policy interventions into Quadrant 4 of Figure 5?"