



The World Bank Should Harness Evidence to Deliver Greater Impact

✦ Eeshani Kandpal, Julia Kaufman, Janeen Madan Keller

Introduction

At a recent CGD event, new World Bank president Ajay Banga called for the institution to deepen its role as a knowledge bank. In his words, “we won’t win the fight ahead of us without taking some risks. The trick is: don’t make the same mistake twice.”

We were pleased to hear this commitment to knowledge generation and use as core to the World Bank’s mission. Simply put, the World Bank cannot enhance its impact, much less lead global development policy, without systematic measurement and evaluation—and knowing what works and what does not across a range of policy options. For instance, which policies increase employment while enhancing gender equality? What are the most effective climate adaptation strategies for low-income countries? Determining the impact of specific programs and policies on development outcomes of interest requires rigorous empirical evidence.

While recent rhetoric—alongside the [Evolution Roadmap](#) and the latest [Development Committee paper](#)—signal a general commitment to this issue, they fall short when it comes to articulating a clear plan for how to incorporate evidence generation and use into the bank’s decision making. The World Bank has a long history of engaging in rigorous assessments of its work, and yet, only [5 percent](#) of its projects have formal impact evaluations.

At CGD, we have spent a lot of time thinking about measuring and improving development policy at the World Bank, among other development organizations. We’ve developed [recommendations](#) for the World Bank to leverage knowledge generation for policy impact as part of CGD’s [Working Group on New Evidence Tools for Policy Impact](#), [argued](#) for greater evidence on what works to [address global challenges](#) like pandemic preparedness and climate, and [researched](#) the effectiveness of concessional climate finance.

In this piece, we share specific ideas on why and how the World Bank's new president should prioritize this agenda as he [advances institution-wide reforms](#). The World Bank's lending function is important and irreplaceable. However, providing countries with an important source of financing and investing in evidence do not have to be in tension. The role of evaluation should shift from an assessment function to one in which evaluation is a central tool of project design and implementation within the World Bank.

Why—and what—is it important to evaluate?

We can measure the net attributable impact of a policy, program, or intervention on specific outcomes of interest by using an approach known as impact evaluation. This type of evaluation is distinct from the important but largely qualitative assessment of program performance conducted by the World Bank's Independent Evaluation Group. The latter typically assesses a project's outcomes against its development objectives, typically at the mid-point and after completion, largely relying on document review and sometimes in-country focus group interviews, but without a clear counterfactual to establish causality.

In contrast, what we describe here is a rigorous approach that is set up at the design stage of the program and uses quantitative data collection and analysis to estimate whether a program or policy leads to an observable change in outcomes using a counterfactual. A counterfactual can be constructed through a randomized control trial, natural experiment, or other empirical tools. Alternatively, it may use related bodies of knowledge to bring together different studies and types of data to answer programmatic questions (see Box 1 for more on how to conceptualize the benefits of impact evaluation). Other kinds of evidence beyond impact evaluation, including qualitative studies, monitoring data, and cross-sectional surveys, provide information on performance, coverage, and implementation considerations and, in turn, shed light on the causal pathways through which a policy or program affects outcomes.

We often focus on examples tied to microeconomic development policy, where impact evaluations are more feasible and popular. But rigorous evidence is also fundamental to good macroeconomic policymaking. There are many examples of World Bank policy claims that are unsupported by rigorous macroeconomic evidence. For instance, the institution's rhetoric on "[billions to trillions](#)" implies multipliers from aid-financed public spending that are two to three orders of magnitude bigger than [what well-identified](#) macroeconomic evidence would support. In a similar vein, calls for the private sector to finance global public goods ignore fundamental economic theory and evidence that the private sector will underinvest in public goods unless incentivized to do so.

Generating such evidence entails a cost and sometimes a time lag, but [recent advances](#) have helped mitigate these obstacles. A typical World Bank project runs over four to seven years. [Adaptive and iterative evaluations](#) with multiple waves of data collection (through low-cost remote surveys, for

BOX 1. THE VALUE-OF-INFORMATION (VOI) APPROACH

The “value-of-information” (VOI) framework seeks to conceptualize the benefits of measurement and assessment. Launched last year, the [final report](#) of CGD’s Working Group on New Evidence Tools for Policy Impact recommended a [VOI approach](#) to impact evaluation funding and practice.

An impact evaluation’s tangible value materializes when it informs resource allocation or implementation decisions, such as preventing expenditure on ineffective programs or informing the scale-up of interventions shown to improve and save lives. Using this evidence produces benefits in the form of increased or faster impact on outcomes or cost savings.

A key VOI ratio is the benefit of the evaluation’s information to the cost of generating the evidence. CGD’s working group report collated [dozens of examples](#) of policy changes due to new information that had tangible impacts on people’s lives, [from development banks, national governments, and NGOs](#). These include [shifting global thinking and practice on reducing user fees for preventive health products](#) like bed nets and deworming medicine; [expanding a cash transfer program](#) to reach more poor households in Ghana; and [informing the timely design of a “low-tech” remote education intervention during COVID-19](#) in Botswana and subsequently other countries, among many more examples. The VOI literature is full of further insights, which organizations like IDinsight are taking forward to [calculate expected VOI for specific projects](#) to help orient research funding towards the most impactful uses.

Through a VOI framework, decision-makers, including funders, consider how much they expect social impact to improve based on new information and seek to fund the highest-value evaluations and studies accordingly. The World Bank can adopt such an approach at scale as part of changes underway to its mandate and operations.

example) and ongoing engagement with implementers can be used to feed into scale-up and real-time course correction decisions for an individual project. Most World Bank projects include a complex set of interventions. Evaluating these interventions can result in important accountability and institutional learning benefits, even if they do not cleanly identify a mechanism relevant for a specific project.

Development finance from multilateral development banks (MDBs) will be relatively marginal in the long-run, especially in middle-income countries, making a pivot towards more systematic knowledge generation and use all the more strategic. Knowledge generation and technical expertise is a clear domain for the World Bank to continue to add and expand its value. Evidence and evaluation should be built into core bank operations, and they can be structured to be agile, of modest cost, and relevant to policymakers, incorporating [advancements](#) in data and design.

Fundamentally, the World Bank and other MDBs must recognize that efforts to drive greater evidence and effectiveness are important levers to advance development effectiveness, value, and impact—not hoops or delays. Some point out that the World Bank [should focus on its lending role](#) and indeed that its knowledge function can [get in the way of investment](#) when researchers are incentivized to focus on novel projects and interventions as part of the “knowledge bank” remit. But knowledge and financing can be symbiotic, so long as World Bank staff and clients tailor evidence approaches to the projects and relevant questions at hand. A “right-fit” [evidence system](#) in which the World Bank helps answer both experimental and observational questions to improve program implementation requires retooled incentives for researchers, discussed more below.

Why should the World Bank’s new president adopt a knowledge-centric approach to development policy?

To date, rigorous empirical evidence from research on World Bank projects has been used to increase the impact of lending programs—albeit on an ad hoc basis. Use of evidence and data can take many forms, including adopting and scaling up—or drawing down and closing—a program; improving program design, targeting, and implementation; and influencing other programs.

Rigorous evidence on how to most effectively spend resources for both traditional [development priorities and global challenges](#) is of paramount importance to the World Bank’s [evolution](#). The potential rate of return is large (see Box 2). The use of data and results could generate significant savings in spending that would have otherwise been mistargeted or ineffective, and can help reduce [trade-offs](#) between funding for poverty reduction and global challenges.

BOX 2. RECENT EXAMPLES FROM THE WORLD BANK’S HEALTH SECTOR OPERATIONS

World Bank projects to improve maternal and child health outcomes through financial incentives directed at providers and consumers of healthcare illustrate the indirect evidence uptake process and potential benefits for informing policies and programs, discussed in Box 1. [Evidence from Cameroon, Nigeria, Rwanda, Zambia, and Zimbabwe](#) shows that providing frontline health facilities with unconditional operating budgets (known as direct facility financing) is as effective as providing them with performance-based financing (PBF)—a financing modality that rewards facilities that meet performance standards in improving healthcare. Direct facility financing results in similar health outcomes as PBF, including increased in-facility births, child immunization, and modern contraceptive prevalence rate, but at significantly lower cost: in [Nigeria](#) and [Zambia](#), direct facility financing resulted in improved outcomes at 50 percent of the cost of PBF programs.

Furthermore, [research shows that demand-side financial incentives](#) that provide subsidies to mothers are about as effective as PBF at increasing aspects of maternal and child health outcomes, but at significantly lower cost: for instance, evidence from Nigeria shows that similar improvements in in-facility births can be achieved through a [\\$14 conditional cash transfer to pregnant women](#) or a [\\$24-per-patient PBF payment](#) to health facilities. Conditional cash transfer programs also have important other benefits, including poverty reduction, increased women's economic empowerment, and human capital gains. Putting these findings into practice would lead to equivalent health gains but at 40–50 percent lower cost, a benefit that far outweighs the costs of conducting the research itself.

However, uptake of research evidence has varied across countries. In a standout case, based on research findings, the Nigerian government pivoted from plans to scale-up PBF, instead scaling up the more cost-effective direct facility financing approach. The World Bank-led evaluation that generated this result cost approximately \$1.5 million. The World Bank operation that scaled-up direct facility financing leveraged \$150 million in investments from the Nigerian government. Assuming that in the absence of the evaluation, the counterfactual project scaled up PBF instead, it would have cost the government of Nigeria \$300 million to achieve the equivalent improvements in health service delivery. This back-of-the-envelope calculation implies up to a 100-fold return on the \$1.5 million investment in the impact evaluation. The gains are potentially so large that they could offset the costs of other evaluations that yield less benefits.

Taking a portfolio-wide approach (in line with the VOI framing discussed in Box 1) to investing in evaluations would help balance benefits and risks: some evaluations will yield such large gains that they will offset investments made in other evaluations within the portfolio that may yield more modest improvements in project design or cost-effectiveness.

But the evidence uptake process is not always as straightforward and linear. In other cases, the PBF approach was scaled up regardless of evaluation findings. A scale-up decision that contradicts evidence signals serious problems with the use of evaluation findings in policymaking and highlights missed opportunities to use evidence to drive greater impact, effectiveness, and value for money. These experiences underscore the importance of incentivizing the use of evidence in directly guiding and informing program and policy design.

A commitment to systematic evidence-informed policymaking is not a small ask. It requires a fundamental shift in the institutional relationship with risk-taking. Such a commitment would also entail independent research being viewed as a core mandate of the World Bank, alongside ensuring institutional structures and incentives to facilitate evidence use. Even measuring the impacts of a substantial share of the World Bank's portfolio is a significant lift that will involve additional research capacity.

The World Bank's Development Economics Vice-Presidency is home to a group of respected and productive researchers. Small numbers of researchers are also embedded in regional Chief Economists' offices. However, all told, the World Bank's fulltime research staff are a small share of its overall staff roster of over 10,000—and the budget allocated to research across the institution is only about \$70 million by one internal [estimate](#).¹ Systematic evidence generation and use would likely require a much larger set of core researchers embedded across the World Bank, not just in one vice presidency and a few Chief Economists' offices— as well systematically leveraging opportunities for productive collaborations outside the World Bank, particularly with researchers, national statistics offices, local universities, and think tanks located in client countries.

Five ideas to foster a culture of transparency and learning at the World Bank

CGD's Working Group on New Evidence Tools for Policy Impact highlighted specific ways for the World Bank to [leverage knowledge generation for policy impact](#), including embedding impact evaluation and related evidence activities across its operational structure. In a [related CGD note](#), we also discussed specific actions to finance and undertake evaluations more systematically across the World Bank's portfolio, especially for global challenges.

The World Bank's overall goal should be to produce and use evidence and evaluations that are independent and rigorous while also connected to the World Bank's policy and lending operations and decisions. How does one engender a culture where task team leaders are routinely engaged in and committed to the evaluation process, no matter the findings? For this to happen, staff incentives must be aligned to foster a culture of transparency and learning. Mr. Banga already understands this: he emphasized the importance of staff accountability and incentives at a recent [CGD event](#).

We have five recommendations for the way forward:

1. Shareholders must place greater value on evidence—backed with commensurate resources

The World Bank's shareholders must demand greater accountability and more rigorous evidence. For instance, IDA deputies should inquire about: what fraction of IDA projects will have rigorous evaluation built into them; whether these projects represent a sizeable share of IDA operations by value; and whether proposed projects are informed by existing relevant evidence on the topic.

1 On page 9, the report states that research accounts for about 7 percent of annual budget for knowledge products. The World Bank spends about a billion on knowledge products each year; .07 of that is approximately 70 million. The report defines research as all work pertaining to "the generation of original ideas and novel methodological tools aimed at increasing the understanding of economic and social issues to inform policy dialogue and influence development thinking. It includes data and tools for strengthening development knowledge. It ranges from country-specific research papers to impact evaluations, to global reports like the World Development Report and Policy Research Reports."

More generally, donors need to be more critical when MDBs report on the “impact” of their work. For example, the significant focus among shareholders on the World Bank’s “corporate scorecard” does not reflect attributable impact of World Bank spending. Beyond scorecards, shareholders should place greater emphasis on rigorous, empirical evidence.

One way to scale and systematize learning from implementation would be to commit funds towards the evaluation of a larger proportion of board-approved World Bank projects. Currently only 5 percent of all World Bank projects are rigorously evaluated, in large part because of how these assessments are funded. Currently, impact evaluations are predominantly funded through a trust fund model dependent on variable donor interest. Having an off-the-top allocation for evaluation that leverages additional IDA funds for all projects could help address such funding limitations.

2. Use multiple evidence-generating approaches to inform individual projects as well as thematic and sectoral areas

Not every project can or should have a large-scale, rigorous impact evaluation attached to it. Still, when they’re conducted, evaluations should be viewed as informing thematic and/or sectoral areas, not just individual projects.

One option is for World Bank teams to implement smaller-scale evaluations or one-off assessments to inform adjustments to design parameters of ongoing projects. For example, a health clinic audit could vary and test different frequencies of audit visits or try a risk-based audit algorithm to better understand the relationship between oversight and the quality of service delivery. A second alternative may be to build evaluations into programmatic support rather than to think of them as informing individual projects. A third idea that has been floated within the World Bank is “batch” project preparation, where researchers would organize workshops with project teams, researchers, and counterparts around a particular sector or theme, to lower the cost of getting the latest evidence embedded into project design and reduce the cost of designing impact evaluations to inform a whole portfolio of projects.

3. Operational staff should be incentivized to follow the evidence

Operational staff are currently rewarded for projects that are approved by the World Bank’s board of directors. This can lead to a fraught relationship between operational staff and evaluators. Instead, could team leaders be rewarded for embedding an assessment into the design and incorporating lessons from the assessment into the project, particularly for scaleup? Equally, on the research and evaluation side, staff incentives must be aligned to help operational teams answer key policy and operational questions that are relevant to them and partner country policymakers, and not just of interest to academic journals. Much more attention should be paid to the quality and impact of research engagement with operations, including for research staff’s professional assessments.

In addition, management must be incentivized to oversee rigorous evaluations that are comprehensive of their portfolios. Importantly, there is also the question of demand from the World Bank's clients. Interest from country partners in using rigorous evidence must also be prioritized and generated, including by designing evaluations that respond to their policy questions and available decision space—several members of CGD's recent working group underscored these [perspectives](#). Further, impact evaluations may require holding back an intervention from a randomly selected subset of the eligible population so that impacts can be rigorously estimated. Clients often balk at the random selection of the group that is left out—even though limited fiscal space and implementation capacity often mean staggered program rollout anyway. To overcome this hurdle, financing could be leveraged creatively—for instance, the World Bank might tie concessional lending to the rigorous evaluation of alternatives to business-as-usual approaches.

Crucially, the World Bank's board and senior management must demand greater accountability from staff. For instance, when evaluative evidence points to the need for significant retooling of a pilot or proposed design, the World Bank board and senior management should demand evidence showing that that retooling is happening.

Through the project review function, the World Bank's board and operations committee can and must lead the institution in adopting a more systematic approach to its relationship with evidence and what it thinks of as “failure.” The board and senior management can empower staff to follow the evidence and, where needed, pause disbursements, or significantly retool a project without fear of professional penalty. This will help build, and more importantly, maintain a culture of partnership and trust between operational and research staff.

To apply these principles in practice, the World Bank's leadership could develop a series of steps that will eventually instill a culture of accountability across the institution. The transition to that culture will undoubtedly be long and messy. During that transition, any of the suggested changes to the evidence generation and review process will only bear fruit if backed by the board and senior management.

4. Research should be quality controlled to protect independence and credibility

Ensuring that assessments are credentialed will guard against the potential capture of the evaluation or assessment process. Research is generally credentialed through publication in a peer-reviewed journal. Where feasible, that could be encouraged and operational staff's contributions to academic publication—not just working papers—could be recognized as part of the World Bank's formal staff evaluation process. We recognize, however, that many—perhaps even most—assessments may not be of interest to academic journals. The World Bank has a peer-review process, which could be empowered, for instance by bringing in external reviewers at both the concept note

and final report stage. Involving researchers at the design stage and moving the concept note review earlier in the process would further increase the entry points for evidence into the policy-making process.

5. Results should be reproducible to increase accountability

We welcome recent efforts by World Bank researchers to invest in the replicability of its analytical work. Such reproducibility should extend to all analytical work conducted across the World Bank. Posting on the World Bank's microdata catalog data and programs used for analysis once an evaluation is approved or published helps promote accountability. Many economics and science journals now have data editors who review the data and programs submitted for all papers accepted by these journals and ensure that results are indeed reproduced in reasonable ways. The World Bank could employ a full-time data editorial staff to similarly ensure that the programs being provided indeed reproduce results presented in reports. These efforts could be tied into investments in data generation and analytical capacity in low- and middle-income settings. The Development Impact Evaluation team's data coproduction model is a great example that could be built upon.

Scaling evidence-informed policy: a win for greater impact at the World Bank

Mr. Banga takes the helm at the World Bank at a crucial inflection point. As the bank advances institution-wide reforms outlined in its [evolution roadmap](#) and operationalizes its [2021 Strategic Framework for Knowledge](#), an evidence-backed approach to policymaking will keep the focus on value-for-money and real-world impact. Project implementation will be where real progress is made.

Additional funds will be needed to systematize evidence generation and use within the Bank's operational structure in the way that we describe here. So, how much of a financial investment are we talking about? Experience suggests that the funds required for a high-quality evaluation would be modest compared to the size of most lending projects; \$1.5–2 million may suffice in many instances. While \$1.5 million may sound like a lot, it is small compared to the size of the projects themselves: a recent IEG report suggests that an average World Bank project might have a budget of in the range of \$120 million.

As an alternative thought experiment, say that the World Bank were to devote 1 percent of IDA loans and grants to impact evaluation. IDA is around \$40 billion per year. Dedicating 1 percent would equal approximately \$400 million per year. While this would be an almost sixfold increase over the \$70 million that is currently allocated towards the bank's research activities, the returns could be sizeable. One percent is not very hard to "earn back" in terms of greater impact—on average, the investment would break even if the effectiveness of spending in the project improved by 1 percent. This is a low bar given the kinds of opportunities to do better that we see in many evaluations, such as the example we spotlight above for a 40–50 percent reduction in cost.

All told, this is a big ask. A shift in the role of evaluation from ad hoc assessment to systematic learning, iteration, and adaptation would entail a large and fundamental cultural shift in the World Bank's functioning. But it is one that would pay off many times over in more effective spending, advancing the World Bank's legitimacy as a leader of multilateralism and development effectiveness.

We're grateful to Aart Kraay, Alan Gelb, Amanda Glassman, and Jed Friedman for several rounds of useful comments and discussions that helped us shape this piece. Thanks are also due to Arianna Legovini, Francisco Ferreira, Dave Evans, Deon Filmer, and Justin Sandefur for feedback on an earlier draft.

EESHANI KANDPAL is a senior fellow at the Center for Global Development.

JULIA KAUFMAN is a policy analyst at the Center for Global Development

JANEEN MADAN KELLER is the deputy director of global health and a policy fellow at the Center for Global Development



www.cgdev.org

This work is made available under the terms of the Creative Commons Attribution-NonCommercial 4.0 license.

