TESTIMONY
OF
PROFESSOR DEAN KARLAN[1]
YALE UNIVERSITY and INNOVATIONS FOR POVERTY ACTION
before the
U.S. HOUSE COMMITTEE ON FINANCIAL SERVICES

"The Multi-Lateral Development Banks:
A Strategy for Generating Increased Return on Investment"

October 9, 2015

In the last quarter century, we have seen huge reductions in poverty around the world. Extreme poverty rates have fallen by more than half,[2] the under-five mortality rate has fallen by more than half,[3] and 91% of kids in developing countries are enrolled in elementary school.[4] But many problems remain, and better evidence on what actually works can help guide us to more effective ways to reduce poverty and improve lives. The multi-lateral development banks (MDBs) have made important inroads in the quest for better policy guided by evidence. But the full potential of their research is yet to be realized. The United States could see a higher return on its MDB investments by encouraging the MDBs to generate evidence on what works in development, build processes to use such evidence most efficiently within their own operations, and disseminate the evidence effectively so that others, such as the United Stated Agency for International Development (USAID) and the Millennium Challenge Corporation (MCC) can benefit. Such "knowledge spillovers" improve the return on investment the U.S. taxpayers receive from their investment in the MDBs.

The MDBs have played a key role in the progress in reducing poverty – but there is still much to do to build stable countries and markets where everyone has the opportunity to flourish. In the coming decades, most of the poor will live in fragile and economically deprived states, making it even more important to U.S. interests and leadership to address extreme poverty. Naturally, resources are limited, so we must figure out the most cost effective approaches, so that limited dollars can have the most impact.

This ought to have bipartisan appeal: those skeptical of aid programs want evidence in order to minimize wasted money on useless programs; those enthused by aid programs want evidence in order to maximize effectiveness with the money spent. Luckily, there is a clear winner: the better the use of evidence to create effective programs and policies, the more leverage US taxpayers can get from our contributions to the MDBs.

---

[1] dean.karlan@yale.edu
[2] For more information, visit
http://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20(July%201).pdf.
[3] For more information, visit http://www.who.int/gho/child_health/mortality/mortality_under_five_text/en/.
[4] For more information, visit
http://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20(July%201).pdf.

So how can we ensure that the MDBs use our precious resources wisely to have the greatest impact on the world? <u>In this memo, I argue that the MDBs are most effective when they are generating "knowledge spillovers," i.e. knowledge that helps other organizations as well and thus influences policy and practice beyond their own operations.</u>

I detail two broad points about the optimal role for evaluation at the MDBs.

First, to date, the MDBs have done their share of high quality research, and there are clear examples of policy-wins that came to be because the MDBs first gathered rigorous evidence on how to deal with specific problems. But their ability to translate research into practice and policy has room for growth. I lay out specific "wins" below, but there are not hundreds to choose from, and we need to position the banks to be more effective in using and applying the most rigorous research throughout their work. To do this, we need to realign incentives and build stronger feedback systems within the MDB processes so that evaluations are used as strategic tools to guide policy reform.
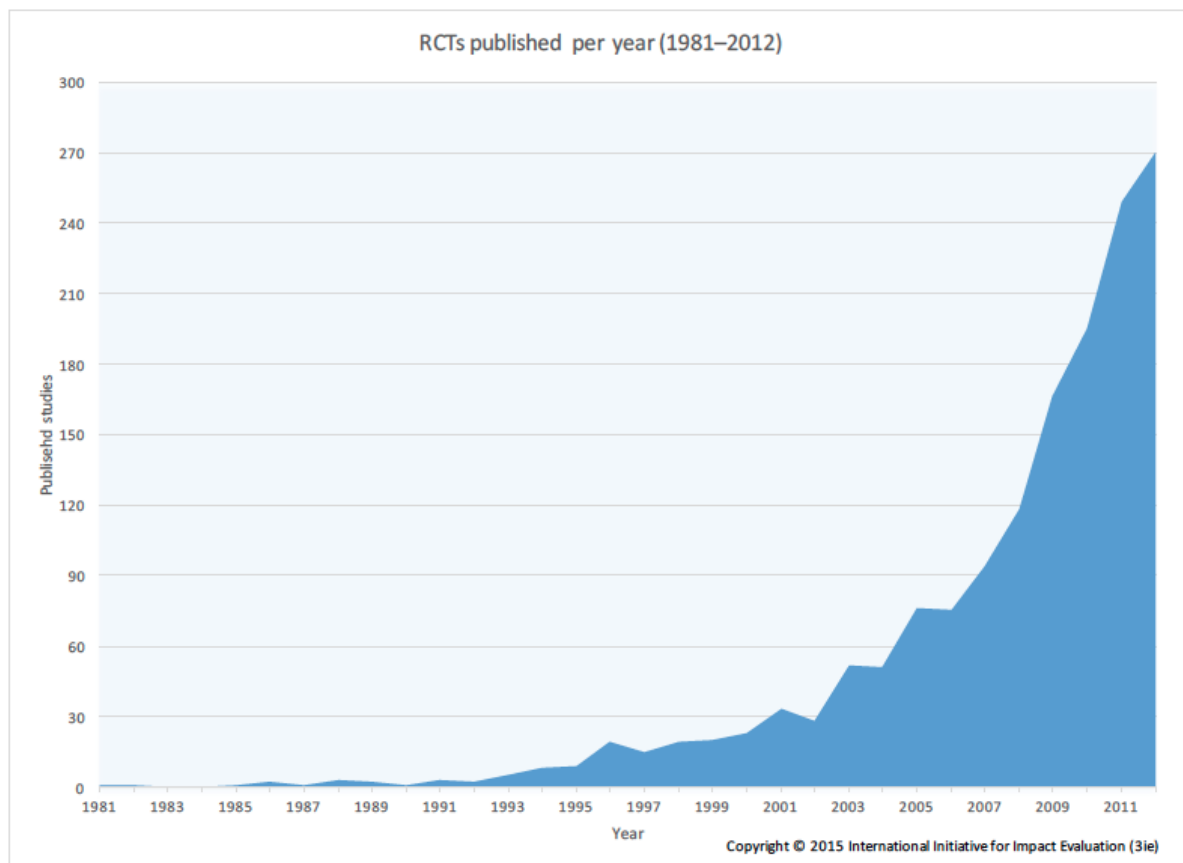
Second, often evaluations have blurred purposes, both trying to evaluate for the purpose of accountability (did the activities under a program or policy take place as expected?) and impact (how did the activities engaged in under a program or policy change the world, compared to how the world would have changed without that program or policy?). Although both questions are critical, they require different toolkits, and not all circumstances call for answers on both accountability and impact.

1. **The Value Added of the MDBs: Translating Knowledge into Policy**

The MDBs are large institutions. But in reality they are quite small relative to the world they aim to influence. Their impact should be measured not by the work they do with their direct funds, but rather with the policies they can influence by producing knowledge for governments around the world. I refer to knowledge for wealthy country governments and their aid programs, such as USAID and MCC, as well as knowledge for less wealthy countries and their programs, such as ministries of health, education, and social protection. This means that they should focus on <u>a key value added: turning good research into policy and practice</u>.

Building this value added of course necessitates conducting rigorous research on what works best at solving poverty problems and <u>what kinds of programs can give us the</u> <u>biggest return on investment</u>.

The MDBs are already doing quite a bit of this research, and in this capacity I have personally worked with multiple colleagues from the MDBs to conduct rigorous research. This rigorous research employs a method called randomized control trials (RCTs), which are changing how development works in the same way that the methodology changed how medicines were evaluated and used. RCTs allow us to measure with confidence and less bias what works and what does not. As you can see from the graph below, the number of RCTs has increased dramatically in development, especially in the last decade, from 19 published RCTs in 1996 and 23 published in 2000, up to 270 published in 2012.

RCTs published per year (1981–2012)

Copyright © 2015 International Initiative for Impact Evaluation (3ie)

But the value added of rigorous evaluation for the world is not just about counting the number of published studies. It is, rather, having strong links from research to policy so that the knowledge created is then used. Strengthening this link for the MDBs requires ensuring that the knowledge created is implemented within the banks' strategies and shared in policy dialogue with other countries.

Examples of Moving from Evidence to Policy

The examples below illustrate how using rigorous research conducted by one or more of the MDBs can lead to more cost-effective programs that impact millions. These examples span from social protection to reduced corruption to improved financial markets for the poor.

1. Conditional cash transfers: Rigorous research led by multilateral development banks has shown that conditional cash transfers are cost-effective methods of transferring funds and generating long-term investment in health and education. Such programs, after clear and rigorous evidence was brought to light, are now being implemented by many governments directly. In the late 1990s, researchers at the International Food Policy Research Institute (IFPRI), with support from both the World Bank and the Inter-American Development Bank (IADB), partnered with the Government of Mexico to test what is now a flagship program: the Progresa-

Opportunidades conditional cash transfer program. The program gives cash to poor mothers on the condition that their children attend school and receive regular medical check-ups. The research also found the program was cost-effective, and the Government of Mexico received other funding to implement the program widely, reaching over 5 million families.[5] Other countries saw that this was both an effective and politically viable program, and dozens of countries[6] are now funding and implementing their own conditional cash transfer programs, reaching more than 20 million families today.[7]

2. <u>Improved credit markets from biometric recording of fingerprints</u>: A program to fingerprint farmers who took out small loans in Malawi was shown effective through research funded by the World Bank. The RCT was conducted by the World Bank's economic research group and showed that fingerprinting bank clients when they took out loans and informing them that the fingerprint would identify them on any future loan applications substantially increased loan repayment among a bank's riskiest borrowers. The technology allowed banks to reliably track clients' payment histories and reduced the costs and risks of lending. The benefits of the technology outweighed the costs by more than two-to-one. Now with support from USAID DIV, Malawi's central bank is supporting the scale-up of a fingerprinting program.[8]

3. <u>Safer public transportation</u>: A campaign which placed stickers in "matatus" (minibuses that are a popular transportation method in Kenya), urging passengers to speak up against poor driving, led to a 50% reduction in road accident claims and a 60% reduction in claims involving injury or death.[9] The evaluation received partial support from the World Bank, and the scale-up is now supported by USAID DIV[10]: with a cost of around US$7 per vehicle, they are now scaling up to 1,000 vehicles per year, Kenya's largest insurance company now requires matatu stickers for coverage, and the government installs stickers during annual vehicle inspections.

4. <u>Community grants for improved health</u>: In Indonesia, World Bank researchers tested a pilot program that incentivized community-based grants intended for health and education services. The incentive stipulated that a portion of the subsequent year's grant would be allocated based on the community's relative performance on health, nutrition, and education indicators, a community-based

[5] Parker, Susan W, and Teruel, Graciela M. 2005. "Randomization and Social Program Evaluation: The Case of Progresa." The Annals of the American Academy of Political and Social Science, Place Randomized Trials: Experimental Tests of Public Policy, 599: 199-219. For more information, visit http://www.jstor.org/stable/25046100?seq=1#page_scan_tab_contents.

[6] For more information, visit http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTSOCIALPROTECTION/EXTSAFETYNETSANDTRANSFERS/0,,contentMDK:22055402~pagePK:148956~piPK:216618~theSitePK:282761,00.html.

[7] For more information, visit https://openknowledge.worldbank.org/bitstream/handle/10986/2597/476030PUB0Cond101Official0Use0Only1.pdf?sequence=1.

[8] Giné, Xavier, Jessica Goldberg, and Dean Yang. 2012. "Credit Market Consequences of Improved Personal Identification: Field Experimental Evidence from Malawi." American Economic Review, 102(6): 2923-54. http://www.poverty-action.org/sites/default/files/publications/credit-market-consequences-improved-personal-identification-field-experimental-evidence-.pdf

[9] Habyarimanaa, James, and Jack, William. 2001. "Heckle and Chide: Results of a randomized road safety intervention in Kenya." Journal of Public Economics, 95(11-12), 1438-1446. For more information, visit https://gui2de.georgetown.edu/projects/zusha.

[10] For more information, visit https://www.usaid.gov/div/matatus.

adaptation of the individual conditional cash transfer model in Mexico. Researchers found that linking community grants to performance improved maternal and child health outcomes. Based on the results of the evaluation, the Government of Indonesia moved exclusively to incentivized grants, and increased its annual funding for the program. With additional funding support from the Australian Government and the World Bank, the program reaches 6.7 million women and children per year. The Millennium Challenge Corporation (MCC) is now funding the Government of Indonesia to scale this program even further.[11]

5. Ultra-poor "graduation" safety net program: A six-country multi-site initiative coordinated by the World Bank's Consultative Group to Assist the Poor, with research conducted in collaboration with Innovations for Poverty Action, found that a multi-faceted and integrated package of services for the ultra-poor can generate sustainable increases in consumption, income, food security, and several other measures of household well-being for those in extreme poverty.[12] This program is now being scaled up by the Ethiopian government (in collaboration with USAID), the Pakistan government and in India, with further work afoot in other countries to bring this program to social protection ministries. CGAP at the World Bank is a driving force in the effort to use this research to shape social safety net policies around the world.

These examples are important illustrations of how research can translate into better programs that deliver on their investments. This wave of development policy based on rigorous evidence is new, as the growth in published studies indicates. It is therefore understandable that we have not yet cracked the code for moving evidence to policy, but this is the direction we need to push to make aid more effective.

2. Accountability vs. Research

Evaluations at the MDBs serve two broad purposes: accountability and impact assessment. It is imperative to be clear about which is called for in specific situations. And not all situations call for both. The MDBs need a more strategic approach for selecting when and how to evaluate different parts of the portfolio. They should adopt a portfolio approach to evaluation, rather than simply employing any one method uniformly. In particular, even though I am an advocate of RCTs in many situations, they certainly should not be advocated in all situations.

So when is accountability called for, and when is impact assessment called for? Let me approach each purpose separately.

[11] Millennium Challenge Corporation. 2011. "Notice of Entering into a Compact with the Republic of Indonesia." The Federal Register: The Daily Journal of the United States Government. For more information, visit https://www.federalregister.gov/articles/2011/11/29/2011-30706/notice-of-entering-into-a-compact-with-the-republic-of-indonesia.

[12] Banerjee, Abhijit, Duflo, Esther, Goldberg, Nathanael, Karlan, Dean, Osei, Robert, Parienté, William, Shapiro, Jeremy, Thuysbaert, Bram, and Udry, Christopher. 2015. "A multifaceted program causes lasting progress for the very poor: Evidence from six countries." Science, 348 (6236), 1260799 For more information, visit http://www.sciencemag.org/content/348/6236/1260799.abstract. Policy bulletin available here: http://www.poverty-action.org/sites/default/files/publications/building-stable-livelihoods-ultra-poor.pdf.

Accountability is an exercise in making sure the program was conducted as it was planned. For a transfer program, this entails examining whether the targeted recipients received the transfer, and the untargeted did not. For a school supplies program, this entails examining whether schools received the supplies that they were supposed to receive, in a timely fashion. For a nutritional support program, this entails measuring whether nutritional supplements reached the intended households.

However, accountability does *not* entail asking how the transfers were used by the households that received them, and how it changed their lives. Accountability does *not* entail asking how the school inputs changed the schools, and thus the students attending the school, their test scores, and later life income. And accountability does *not* entail asking how the nutritional supplements have helped improve body-mass index of the children in a household.

One often finds that evaluations aiming for accountability stretch into measuring *outcomes* when they ought not. Let's take the last example, the nutritional supplements. Suppose children receiving the nutritional supplements did increase their body-mass index in the three months following enrollment in the program. Would this constitute evidence of impact? No. Children grow anyhow. The question one must ask is whether they grew more than they would have had they not received the nutritional supplements. This is a much tougher question to ask, and enters into the space of impact assessments, where RCTs have become commonplace.

But suppose we already know that the nutritional supplements "work." They have the right content, and are tasty. There may be no need to know their impact, just like we do not likely need another RCT of penicillin. It may also be difficult, due to contextual realities, to conduct a proper impact assessment of a specific program. In either case, it would not be prudent to invest resources to measure the impact of the program. But an accountability exercise ought to be done, merely as a check on the organization, to make sure it is adhering to its plans. Such accountability exercises are radically cheaper than a full impact assessment, and ideally are at least partly generated through data collected for management purposes already.

When the MDBs evaluate programs, they must be clear about the purpose of doing so – rigorous evaluation is for determining impact and why, but not every project ought to have a rigorous impact evaluation. That would constitute too much money being spent on research. Rather, impact evaluations should be deployed strategically, not applied universally. In particular, once there is evidence on the impact of specific policies in well specified contexts, further research on impact is unnecessary, and those funds are better used filling in other knowledge gaps, or providing direct services. Reverting back to the analogy in medicine to make this point: once a particular treatment is known to work, one does not continue to conduct rigorous research testing its efficacy.

Research on the other hand should answer the broader question of impact: how has the world changed because of this policy, compared to how it would have changed otherwise,

why, and will the results replicate in other contexts? There are three parts to this question, and it is important to deal with each.

The first part asks the basic impact question, but critically includes the causal argument that we must know how the world changed compared to a counterfactual, i.e., what would have happened without the program. RCTs became popular in development specifically because of their ability to estimate this counterfactual with less bias.

However, it is also critical to answer the "why," because understanding why something works is critical for taking ideas to scale. Understanding the "why" is akin to asking whether we have a theory that explains the observed results, and ideally this theory also predicts in what circumstances we can expect similar results. This helps us to leverage research from one context to make well-informed decisions about whether the same policy or program is also likely to be effective in a new context.

The third part, will the results replicate, is a reality check on the first two. Many evaluations may claim to have a clear understanding of the impact, and even the reason why the program worked, but putting such theories and claims to the rigorous test of a replication in another context is critical for taking ideas to scale.

To illustrate research that meets these aspirations, I highlight the fifth example above, the "graduation" program for building sustainable income and well-being for those in extreme poverty, particularly because this project was born out of a collaboration led by an MDB. The program puts forward a simple but powerful concept: the problem with being ultra-poor is not any one constraint, but several interdependent constraints, and attacking poverty with an integrated program may be necessary to generating long term improvements. The "graduation" program includes a grant of income-generating assets (e.g., often livestock), training, coaching, access to formal sector savings, health services and short-term food transfers. The program typically targets those in extreme poverty, living on less than $1.25 a day.

This work began from a partnership between the World Bank's Consultative Group to Assist the Poor (CGAP) and the Ford Foundation, who then enlisted Innovations for Poverty Action to conduct six randomized control trials of this program in six countries. I was a lead researcher on this work. The goal here was two-fold: first, to learn whether this integrated package of services was sufficient to generate a "big push" that helped elevate income and consumption sustainably. Second, as discussed above, by conducting this research in six different countries, we learn how replicable the results are across a multitude of implementing organizations and country contexts.

The research, published in *Science* (available [here and policy bulletin available here](#)), followed 21,000 of the world's poorest people for three years, and found that this approach led to large and lasting impacts on the ultra-poor's standard of living and boosted livelihoods, income, and health.

Now that we know the approach works across contexts, the MDBs are already active on two key fronts: bringing this evidence to policy, and leading further research to better understand why this program works (with an eye towards guiding improvements in how it is run and how to make it even more cost effective, particularly when implemented at scale). On the policy front, the government of Ethiopia, which has received support from USAID on this project, expects to scale this model to 3 million people, and the model is already being scaled in Pakistan and India. On the research front, further research is underway in Ghana, Uganda, and the Philippines to help determine if all of the components of this program are necessary. Such knowledge helps understand *why* the program works, and with such knowledge one can build an even more cost-effective program at scale.

Research conducted by the MDBs and institutional policies that tackle these three roles for research –understanding is there impact, why, and does it replicate – can have transformative effects, leveraging a relatively modest research budget into larger impact through its influence on how other MDBs, government, and bilaterals allocate their funding.

3. Conclusion

There are many steps the MDBs can take to facilitate further evaluation that serves this forward-looking strategic purpose, i.e. evaluation intended to guide future decisions. The MDBs have made huge inroads, as discussed above, in the quality of evaluation produced over the past two decades. Improvement naturally should be an ongoing effort, and establishing clear institutional policies can help guide them towards even higher quality standards.

Specific examples of policies and processes to improve MDB policy through evidence

1. Peer reviews, both external and internal, can help provide useful feedback for MDB evaluators.
2. Engagement with peer reviews, external and internal, can help make sure evaluations are addressing key knowledge gaps and not merely serving accountability purposes (as discussed above).
3. More evaluations should take on a multi-site approach, coordinating on specifics of measurement as well as intervention design, so that the world can learn more systematically what works and what does not, rather than accumulate a series of seemingly-related studies. This is akin to mimicking the approach in the fifth example provided above, coordinated initially by CGAP at the World Bank, on the ultra-poor "graduation" safety net program.
4. More systematic engagement of the evaluators at the MDB research departments in follow-on policy decisions and technical assistance after the evaluation, so that the learnings from the evaluation are best used in the next phase of the government policy.
5. To improve the use of evidence to guide decisions, a formal process should require close collaboration between experts in evaluation who are deeply entrenched in the existing knowledge on a topic, and those in operations responsible for implementation. Such integration is already happening at MDBs, and more to

encourage further and tighter collaboration would prove fruitful for improving the return on investment in the MDBs.

Broad policy recommendations

To summarize, this memo lays out the following broad policy recommendations for the committee to consider. The overall goal underlying each point is simple: leverage US government funding to the MDBs by helping create more knowledge, and helping that knowledge get into the hands of critical actors such as government and bilateral donor agencies.

1. Take a portfolio approach to evaluation, encouraging MDBs to *not* overinvest in impact assessments. Impact assessments should only be conducted when knowledge can be generated with an eye towards benefits for others; other situations may be best off with merely accountability exercises, which document that the planned activities were properly implemented.
2. Ensure that the internal organizational structure of MDBs enables the facilitation of knowledge from research to policy (see four examples above).
3. Fund programs backed by evidence – there is already ample rigorous evidence out there showing programs that not only work, but that are cost effective. I named only a handful. But the more we demand programs can demonstrate their impact, the more likely we are to achieve impact and use money effectively.