MILLENNIUM
CHALLENGE CORPORATION
UNITED STATES OF AMERICA

Transcript
October 29, 2008 | www.mcc.gov

Millennium Challenge Corporation Hosts a Public Outreach Meeting

Rigorous Evidence:
Key to Progress Against World Poverty?

Speakers

Jon Baron,
Executive Director,
Coalition for Evidence-Based Policy

Franck Wiebe,
Chief Economist,
Millennium Challenge Corporation

Ruth Levine,
Vice President for Programs and Operations,
Center for Global Development

Mark Lopes,
Senior Policy Advisor,
Office of Senator Menendez

Rachel Glennerster,
Executive Director,
Abdul Latif Jameel Poverty Action Lab

Ron Haskins,
Senior Fellow,
Brookings Institution

Dan Levy,
Lecturer in Public Policy,
Harvard's Kennedy School of Government

Transcript

BARON: OK. We're going to go ahead and get started.

This forum is sponsored by the Coalition for Evidence-Based Policy in collaboration with the Millennium Challenge Corporation.

My name is Jon Baron. I'm the executive director of the coalition, and I'm going to say just a word or two of introduction about the coalition and then also a few introductory thoughts to this meeting for your consideration on evidence-based approaches to development assistance.

The agenda for this meeting is the first thing in your packet. As you'll note, it includes a good amount of time for question and answer and open discussion, and we'd be very interested in any thoughts or suggestions that you may have on the ideas that are presented today. There's a lot of time built in for discussion.

With that, let me introduce the coalition. We are a nonprofit, nonpartisan organization. Our mission, in a sentence, is to increase government effectiveness through rigorous evidence about what works in domestic social policy in the United States and also in development policy in other areas.

We have a bipartisan board that includes, for example, David Elwood, who's the dean of the Kennedy School at Harvard University; Bob Solow, Nobel Laureate in economics at MIT; Dan Levy, who's here, going to be speaking later today; and Ron Haskins, who's also on our board and is going to be one of the speakers; and others. I'll show you the board in a second.

But we have worked with Congress, the federal agencies, the Office and Management and Budget to advance evidence-based reforms in government programs, and our work has helped advance some key reforms.

Just to give you a couple illustrative examples, we worked with Congress in a bill that was just signed into law to facilitate the reentry of prisoners into the community. In that bill that was just enacted, we worked with Congress to get a 2 percent set aside for rigorous, preferably randomized evaluations to build evidence about which reentry strategies are truly effective and which are not.

We also worked closely with OMB and Congress last year to get enacted into law a new evidence-based home visitation program at HHS designed to scale up research-proven programs of home visitation for poor, low-income women in the United States. And there are other examples.

We are not affiliated with any programs or program models, and so we serve as a neutral independent source of expertise to Congress, the federal agencies and others on evidence-based programs. Our work is funded independently by the MacArthur Foundation, the William T. Grant Foundation and the Clark Foundation.

And here's our board of advisors.

Turning now to the subject of this meeting, just a few brief introductory remarks to get the ball rolling.

In development assistance, the problem that evidence-based policy seeks to address is this: That development agencies, developing country governments and other organizations spend tens of billions of dollars each year to help the world's poor, but very little scientifically valid evidence exists about which strategies they fund are truly effective in reducing poverty and which are not.

That was the central finding of the influential evaluation gap report of 2006, which did a comprehensive review of the evaluation literature in a number of different areas in development policy and found, and I quote, "that for most types of programs a body of scientific evidence about effectiveness is lacking. For almost all projects currently in operation or in the pipeline, virtually no credible information will be generated about program impact." That was the latest and probably the most comprehensive review, but earlier World Bank reviews in specific areas within development policy have reached similar conclusions.

Now, recently there has been a recognition of this problem at the World Bank and the Millennium Challenge Corporation and elsewhere and some important initiatives to begin to address the evaluation gap, which are going to be discussed at this meeting.

But we would suggest that rigorous evaluation can succeed in building a body of scientifically valid, actionable evidence about what works, that can spark rapid progress in development policy in the same way that rigorous evidence has sparked rapid progress in medicine over the past half century.

And the reason we believe it's possible is this: It can be illustrated with two quick examples -- one from the United States domestic policy and one from the developing world. And there's a quiz at the end of this example, so stay alert.

In 1980, the Department of Labor in the United States launched a demonstration program in which they provided vouchers -- in which they provided a subsidy to employers to hire low-income workers. And the way it worked is that they gave each low-income worker -- usually they were welfare recipients -- a voucher that that worker could hand to a prospective employer, and if the employer hired that worker, they could cash the voucher. It was a sizable voucher.

The demonstration was set up as a randomized experiment where some low-income workers got the voucher and a control group did not. At the end of the experiment, the workers in the control group who didn't get the voucher were 60 percent more likely to have landed a job than the workers in the treatment group. OK. So the control group had far outperformed the treatment group.

Does anyone want to guess why that occurred? Yes?

AUDIENCE: A voucher would make someone look unattractive.

BARON: Yes, that's correct.

The voucher had stigmatized the workers in the eyes of the potential employers, and it actually backfired.

As a second example -- so these are -- as a second example, this one from the developing world but something that's been found not effective, there was a study that was done in Indonesia about five years ago where researchers conducted a rigorous evaluation of road construction projects to determine grassroots monitoring of the road projects could succeed in reducing corruption and waste.

This study was set up as a randomized evaluation that randomly assigned over 600 villages in Indonesia that were received funds for these road projects to either a group in which the villagers were invited to village meetings where project officials accounted for how they were spending the funds and where villagers were also encouraged to make anonymous reports of any corruption or waste that they knew of. Or the villagers were randomly assigned to a control group that did not have this grassroots monitoring.

At the completion of the project, the study found no effect on the amount of corruption and waste in the projects. The amount of corruption and waste was measured independently by a team of engineers who estimated how much the project should have cost, based on the materials that were used and so on, compared to what was actually charged. And this intervention, the grassroots participation, had no effect on that measure.

So part of what rigorous evidence can offer, we would suggest, is a way to identify those ideas that you would think ought to work, which sound good in theory, which may be backed by expert opinion, but which when tested in the field with a rigorous evaluation turn out not to work for reasons that had not been anticipated. And, as I'll mention in a moment, there are many, many, many ideas in this world that are like that from all different areas of policy.

But at the more positive end of the story, rigorous evaluations also have the potential to identify a few interventions, a few strategies that are highly effective and can make meaningful improvements in people's life outcomes again, one example from the United States, probably the leading example from domestic social policy of a research-proven program with large effects, and one example from developing countries.

Domestic policy, this is a program in the United States called the Nurse-Family Partnership. Very briefly, it is a program of nurse visitation for women who are poor, pregnant and mostly single. These are usually women in their late teens. The nurse visits the woman during her pregnancy and the first two years of her child's life and teaches her basic parenting, nutrition, not to smoke, not to drink during pregnancy and, if she's interested, how to practice birth control afterwards so she doesn't immediately have another child.

This has been evaluated in three large randomized controlled trials with long-term follow-up and been found in all three cases and populations to be highly effective. In one of the studies, which was an age 15 follow-up, when the children reached age 15, this program produced 40 to 70 percent reductions in criminal and child abuse and neglect and in criminal arrests and convictions of the children of those mothers. So a large effect in long-term follow-up, replicated across three different studies. That's domestic policy.

In the developing world, this example also from Indonesia, again road construction projects. This study randomly assigned over 600 villages to a group where it was announced that all road projects in those villages would be audited by the government versus a control group where it was announced that the typical auditing would be done where only about 4 percent of projects were audited.

At the project's completion, the study found that the intensive auditing reduced project corruption and waste by about 30 percent, again, as assessed by independent engineers, and, importantly, the auditing more than paid for itself in reduced costs of the projects.

So part of what rigorous evaluation can offer is a way to identify a few ideas that concrete interventions like this that can make meaningful improvements in important societal outcomes and in people's lives. Between these two extremes what's been shown effective and what's been shown ineffective -- what I discussed earlier -- probably lies, what, 98 percent of everything else that governments and private philanthropies and others, development organizations, fund in this world where we really don't have a strong idea of what's going on.

Let me ask you to hold questions till the end of the session.

So those are a couple thoughts on, sort of, a general rationale for evidence-based approaches in this area.

Let me just offer two thoughts on evaluation strategy for your consideration.

Our thoughts on evaluation strategy are based on two observations. One -- and I'm not going to go into details here -- is the notion that how one measures whether a project is effective, that is the specific study design, we would suggest, can be extremely important. In particular, the strong evidence to suggest that random assignment studies, like the kind I talked about earlier, where they're well implemented, are the strongest way of telling whether a project works or not. They're considered the gold standard in medicine and education and welfare policy and other fields for determining what works.

And the reason is fairly straightforward: If you randomly assign -- it's randomization. If you randomly assign a large number of individuals or villages to two different groups, that assures that your groups are equivalent in all factors except for one: The program group gets the program and the control group does not. And so any difference in outcomes between the two groups over time can confidently be attributed to the program and not to other factors, which were controlled for.

Where random assignment is not possible -- and in a number of cases it is not feasible -- there's good evidence to suggest that comparison group studies where the program group is compared to highly similar individuals or villages before the program is the second best alternative.

But the other types of studies that fall below that, where you're comparing program participants to groups that are not equivalent, for example, these kinds of studies -- I'm not going to go into detail -- but often produce erroneous conclusions. They can be useful in generating hypotheses about what works, but there's a lot of evidence from many different fields that they often produce the wrong answer about whether something really worked or not.

So that's the first point about the importance of the type of study that's done in determining program effectiveness.

The second point that I'd like to leave -- and my last point, that I'd like to leave you with is this: That much of the conventional wisdom about what works in development policy, domestic social policy, medicine and virtually every other field is probably wrong. Specifically, much of what is thought to work probably does not work or has weak effects. Findings of true effectiveness, like the Nurse-Family Partnership that I discussed earlier, like the government auditing of projects, those examples exist but they tend to be the exception. And that pattern occurs across many different fields.

Let me just say word here about medicine. These are examples of the conventional wisdom that the medical establishment thought were true for a long period of time that were overturned in gold standard random assignment studies. One example was in the news about two or three months ago. Doctors have been telling their diabetic patients for 50 years to make intensive efforts to lower your blood sugar to normal levels.

When that strategy was actually tested in a randomized, controlled trial, intensive lowering of blood sugar versus the usual less intensive efforts to lower your blood sugar, one of the studies, one of the approaches actually increased the risk of death -- increased the rate of death in the treatment group compared to the control group, and the study had to be stopped. The other one found no difference, which was a shocked, and it overturned, potentially, half a century of conventional medical wisdom on how to treat Type II diabetes.

There are many other examples. Hormone replacement therapy for post-menopausal women is another example where the conventional wisdom was turned on its head by a large, randomized trial, one that was actually in the news yesterday. I have the New York Times article here. We all heard about the antioxidants, the great anti-aging properties and anti-cancer properties. Almost all of them seem to be falling by the wayside now that they've been tested in large, randomized clinical trials, the latest one being vitamin E, which was reported on yesterday having no effect on and preventing prostate cancer.

And the last one I'll mention here, because it leads into my next point, is the promising AIDS vaccine that was tested in a large, randomized clinical trial in Africa and was found, despite earlier studies suggesting it would be effective, was found not to work. It actually had a small backfiring effect, which was very disappointing.

But I want to mention something here. The head of the AIDS Vaccine Coalition said this after learning of these study results: That this is an important milestone in many respects. This is the way new products get developed. Lots of things don't work, and we're on our way to finding something that does. I just want to repeat that: This is the way -- lots of things don't work, and we're on the way to finding something that does.

We recommend that policymakers take a similar approach in development policy and domestic social policy for the following reason: Lots of things in this area don't work either, and these are just -- I'm not going to go into details on any of these, but these are examples from domestic policy of well-intentioned, sometimes well-designed interventions that when tested turned out either not to work in a randomized control trial or were to found have weak effects --most home visitation programs, other than the Nurse-Family Partnership for low-income families, the big after school program at the Department of Education, educational vouchers, leading educational software, a prize-winning software for teaching reading and math, our nation's largest substance abuse prevention program. Most of the big federal job training programs have been found to have weak effects or no effects.

The bottom line is that much of what is thought to work in almost all other fields probably does not work. Findings of true effectiveness exist. They tend to be the exception. And so a central suggestion that we make on evaluation strategy is not -- I repeat -- not to try to rigorously evaluate everything or even most things but to focus scarce evaluation resources on the most promising interventions in order to find a few things that do work in a meaningful way and then scaling them up.

And with that, let me introduce our panel. We have Franck Wiebe, who is the chief economist for the Millennium Challenge Corporation. He's one of our speakers, our first speaker. After him will be Ruth Levine, who's with the Center for Global Development and Mark Lopes is going to be a discussant on the panel responding to the presentations. He is senior policy advisor to Senator Menendez who chairs the International Development Subcommittee of the Senate Foreign Relations Committee.

And with that, please join me in welcoming Franck.

(APPLAUSE)

WIEBE: Thank you, Jon.

And with that dose of pessimism, I can't help but be a bright light of optimism.

Actually, I appreciate that background, Jon, and I really do hope that by talking about MCC, we'll be able to put perhaps a more optimistic spin on, first, what's being done in foreign assistance and what our opportunities are.

Let me also say that for those in the back there are chairs up front here, and so if you want to, while I am getting set up, if you want to come up and have a seat. I'm tempted -- I know, but you don't need to put this in the transcript, I'm just going to make a joke.

(LAUGHTER)

I'm tempted, seeing my colleagues in the back, to say that these aren't necessarily latecomers but are actually selecting proximity to the lunch, but, of course, I would have to do a randomized sampling to see whether that were true or not. But, again, if you guys would like to come forward, and for those who want to stay in the back, please, save me a sandwich.

Again, thank you, Jon, for that background and the introduction, and let me express appreciation to you and the council for organizing this conference. It's clear that the council is an institution that lives up to its name.

When we first started talking about this event, we had a number of things in mind. One of the things that we wanted to do at the MCC was make sure we reached out to a broader audience that might not be as familiar as some of our normal stakeholders in how foreign assistance is currently conceptualized, some of the cutting-edge practices, but also those who are really interested in the issue of accountability in government, because at MCC this is one of our core principles.

And so this discussion of impact evaluation is not just about foreign assistance but it also touches on the issue of accountability in government. And so we wanted to make sure that the audience was broader than some of our events that are focused specifically on the foreign assistance community.

One of the things that I've also noticed from Jon's presentation is, it's very clear that it's much more interesting to talk about the actual impact evaluations underway -- is this not going forward, Jon?

BARON: Press the lower button.

WIEBE: Still not.

One of the things that's clear from Jon's presentation is, it's much more interesting to actually talk about impact evaluations, the stories, the lessons that we've learned -- am I pointing it the wrong direction? It's like when you're watching TV and you've got the clicker in the wrong direction.

And so my challenge is a little bit different. I'm not going to be talking about specific impact evaluations and so will miss, in some sense, the benefit of talking about interesting case studies and what we've learned. Like I said, Jon's presentation has already touched on that -- Oh, I see, so go ahead to the next slide, and you can go to the next one as well.

My task is a little bit different. I need to talk about MCC's institutional approach to impact evaluation, how it fits into our broader foreign assistance model. And I think that I would be safe in claiming that that's going to be more of a challenge, that is a challenge to make it interesting to the audience. I will try to do that.

I will also try to keep it brief. I was in Chicago a couple of weeks ago speaking with a group of students at the Harris School of Government there, and my one-hour session actually went to almost three hours and didn't end until I came out and said, "Look, we don't have any more paid internships," and then everybody left.

(LAUGHTER)

So let me say that up front. But I'll try to keep my comments brief.

I want to talk about three things today. The first is MCC's model, because I think it's important to get a baseline understanding of what it is that MCC was set up to do. And the second thing then is, how our investment and impact evaluations fits into that broader focus on results. The third thing that I want to do is to do a very brief demonstration of a new Web site interface that we have that's recently been put online, and I want to encourage people in the audience here who are interested in this to go online and learn more about our impact evaluations that are underway. So that's my -- I'm going in the wrong direction. OK.

So let me start with MCC's model. MCC was established four years ago, and what I've learned is I have to go back to my old slide presentations. I can't really any longer say, MCC is a new and exciting foreign assistance agency. I can say, though, that it still is an exciting foreign assistance agency. And it was created as a new agency to do things a bit differently.

And I want to emphasize two things in my talk about our model. One is MCC's explicit focus on reducing poverty. In our cards, it says, "reducing poverty through growth." On much of our letterhead, we had this as our motto, and it's something that we take very seriously. If you will, it's a single objective that MCC looks at and concentrates on in a way that, to be honest, is often not found in other foreign assistance agencies. And, in fact, I think it's fair for me to say that because many of those agencies would defend what they would consider a broader, more complicated story about development.

What I think is important is MCC's focus on poverty actually allows it to do things a little bit differently.

The other things that's important to note is that when MCC was created four years ago, it was built explicitly with a number of best practice principles in mind that were, perhaps, not new, but I think it's fair to say that at that time were newly described as generally accepted within the international foreign assistance community.

I want to talk about the -- oops, not that fast. I need to go back.

Jon, can you help me? How do I go back?

BARON: The other button.

WIEBE: My own technological shortcomings. Oh, there we go. OK.

So let me talk about four fundamental principles of the Millennium Challenge Corporation. The first one is that growth matters, and this is an important distinction. It was an explicit philosophical statement that the core element of any poverty reduction strategy has to recognize the role of economic growth. And, again, this is something which is broadly now accepted within the development community, that countries that grow faster see their poverty decline faster. Countries that don't grow very rarely, if ever, experience significant poverty reduction.

This focus on growth is a core philosophical statement from MCC from its very outset, and I think that -- I'll return to how this is embodied practically in the institution. It turns out this isn't important just as a philosophy, it's important practically as well. It has important practical implications.

The second is that good governance matters, and that is that the effectiveness of foreign assistance is determined, at least partly, by the policy practices of the recipient countries. MCC uses a series of 17 indicators as a means of determining eligibility. MCC doesn't work with every country; rather, we work with countries that we identify as potentially strong partners who have demonstrated through their own practices that they have a basis for using aid effectively from MCC.

The third principle is that country ownership matters, and this is something which in the aid effectiveness literature is sometimes also described as alignment with country principles, but MCC takes a broader view of it as well. First, we ask countries to develop the proposals themselves. Once we establish that countries are eligible for working with MCC, we ask them to submit the proposal. So it's not MCC staff developing a proposal; it's something that emerges from our country partners.

The second aspect of this country ownership is that country partners are responsible for implementing the program as well.

MCC is a very small agency. Right now, I think we have about 280 staff. In our 18 countries where we have programs, we have no more than two or three MCC staff present. And you say, "How can you implement a program that's hundreds of millions of dollars with two or three staff?" The answer is, "We don't." MCC doesn't implement the program; rather, our country counterparts are responsible for implementing the program that they have submitted to us.

And then the final core principle is that results matter. And for MCC, we take a very unique approach to this, I think, a very systematic approach in a number of ways. The first is that we are looking for objective measures of impact and the second is that when we talk about this, we not only collect information about how the programs will be used but that that information is used in a very transparent and public way in the decision-making process.

I'm going to talk a little bit more explicitly, but, in fact, the focus of my comments today are really on this fourth bullet, and that is how MCC's model embodies its emphasis on results.

So, first of all, I think it's important to take a step back -- even though my time is already running short -- step back and start and say, "What is it exactly we're talking about aid effectiveness." We have to, first of all, come to a common understanding of what we mean by that.

So let me put forward one definition of aid effectiveness. You can read along with me. "Aid effectiveness is the effectiveness of development aid in achieving economic development or development targets."

You know, this is a horrible definition. I usually look for my audiences to start laughing when they see that. Maybe you already had the sandwiches already in you, but you look at a definition -- now, I'll admit, Wikipedia shouldn't necessarily be the source of technical definitions, right, source of all knowledge. On the other hand, it's surprising how many of us use it for a lot of things, and this is how they define aid effectiveness. And to be honest, it's not a bad representation the way the term is used in the development community. And this is a part of the problem. That's the point I want to get across.

In fact, if you look -- as you know, the hyperlinks, obviously, you can't click on it now, but if you clicked on "economic development," what it would tell you is -- to me, that's at least, kind of a, partly sensible term -- but you click on that and you say, "But don't go too fast. Economic development is a term itself which can't really be defined. It certainly isn't economic growth. It means all kinds of things."

So we have a term here which uses the same words in the definition, and then the one word that's new, it says, "And you can't define that either, at least not easily. It's a very complicated process."

Again, my point is that even the aid effectiveness literature has a hard time grappling with what do we mean by aid effectiveness.

And this is where I'm going back to the philosophical point about reducing poverty through growth. Because MCC focuses on growth and income, it actually allows MCC to break out of that mold. We have an infometric that's directly linked with measured poverty, as you're familiar, and now the World Bank says it's $1.25, not $1.08. But this notion of using an infometric, even as an imperfect proxy for welfare, it's commonly accepted and regularly used throughout the world. MCC uses that infometric as well to say, "This is the poverty that we're dealing with."

What that does is it allows us to do project appraisal across sectors where we can use the same objective and compare potential results across the different proposed investments. It allows MCC to break out of what one of my colleagues made me say is some positive income trap. He's here. I like to use the term, "non-zero," because if you talk to a lot of aid practitioners and you say, "How effective was your program, " they can tell you what they accomplished. And you say, "Well, was that enough to justify the cost?" Well, it's something. In other words, every aid project can claim some positive impact, some non-zero impact. That doesn't really get us very far. It certainly doesn't allow us to answer the question, was the aid effective given the cost that was invested in it? This model allows MCC to break that trap.

I'm going to start then and very quickly run through what I like to call the cold chain of the aid effectiveness at MCC. Cold chain is a term that comes from the public health literature and refers to vaccines, and, as I'm sure all of you are familiar, you know, vaccines are produced in some place in Switzerland, and they're actually a living substance. But we don't give them -- well, sorry, you know what I'm saying.

But if we want to deliver them in a village in Africa, for them to be effective, those vaccines have to be kept cold from Switzerland all the way to that village and the point of delivery. And so there's big literature within public health about how you maintain the cold chain. I'm trying of establish -- probably having very little luck -- establish the same use of this term in aid effectiveness literature by suggesting that if we want to be effective what we have to be able to do is talk about aid and our results and maintain that link from before we invest to after we implement to know whether in fact we're having the impact that we claimed we would have.

MCC does this in three or four different steps. The first main one is what we call pre-investment analyses. The first half of that is what we talk about as a constraints analysis but perhaps is more commonly known in the broader development literature as a growth diagnostic coming from Roderick Housen and Glastow (ph) at the Kennedy School. But, basically, it's a way of asking countries before they submit a proposal to us to think about growth and to identify what the bottlenecks are in their economy. So before they ask us to build a road or to build a clinic or to build a school, we ask them to come to us with a description of the problems that they have that are impeding economic growth.

As they think through that, the next step is they submit a proposal to us, and already in that proposal, one of the things that we ask them to do is to conduct a benefit-cost analysis of each activity to determine whether the expected results justify the cost of the investment.

I like to talk about benefit-cost analysis as a pre-investment impact assessment, because, basically, what you're saying is, "I need to think through what the program is going to do, who is going to benefit from it, how much they're going to benefit from it and then compare it to the cost. Again, this allows us to use a monetary metric of benefit, it allows us to incorporate the policy reforms, the institutional changes that the countries say to us that they need, and then it allows us to compare proposed activities across sectors and, ultimately, at some point, one can compare across countries as well.

Then once programs are approved, we go into -- each project activity includes a monitoring and evaluation plan -- every project has an M&E plan that includes a baseline survey before the program starts. One of the big problems in evaluation in the foreign assistance community is often it's not considered a priority until after the program's already been implemented, which makes evaluations very difficult. MCC begins by doing baseline surveys for its activities before any implementation starts, and that then allows us to monitor implementation against expectations.

And the last step of this cold chain is to do rigorous impact evaluations as appropriate. So where projects are -- where the information about projects beforehand is less concrete or where there's a lot of opportunity for learning, MCC identifies an opportunity for impact evaluations.

Let me say that this whole chain is described in the paper that you may have seen on the table, "Aid Effectiveness: Putting Results at the Forefront." I haven't read it yet, but I've heard it's very good.

Let me also say that we're going to be talking about this last step for the rest of my few minutes, and for that I wanted to point out that we've also distributed a paper by Abaji Banaje (ph) from the Poverty Action Lab at MIT, and if you're not familiar, that essay comes from this book -- Jon, are you going to make me stop now?

BARON: No.

WIEBE: Oh, good. One colleague of mine saw this and he said, "Wow, is that really all it takes, making aid work, one book?" Well, maybe not all of that. I mean, this is, again, one part of our story, I think, but I would encourage you, first of all, to read the essay, if you've seen it. There's also a response from Ruth Levine on my panel -- on our panel here. But for those of you who haven't seen the book, I would encourage you, it's an excellent read.

So now, in terms of impact evaluations at MCC, the point being that for many activities simply monitoring performance against your expectations is not enough. You're not able to see what the true impact is. And this Jon has already gone over, so I'll go over this very quickly.

Impact evaluations allow to provide assessment of the projects based on a counterfactual. That's really one of the core issues of an impact evaluation. What you're doing is you're actually measuring what happened to people who weren't in the program and where possible to use randomization to strengthen the rigor of that.

Impact evaluations provide highly credible evidence, and we use this in our programs in a number of ways for testing implementation approaches. We can find which ways of doing a program work better than others. It allows you to pilot activities and not spend money until you have confirmed that the program works the way that you had hoped and that it has the effect that you had expected. And then it also allows future funding decisions in terms of thinking about not just this program within MCC but also future money spent by country governments or other donors.

MCC does this for roughly 20 percent of its activities. And as I said, we do baseline surveys for all of our activities, we monitor all activities, and we do post-implementation assessments of all activities. And already that and our ability to do a quantitative assessment for everything that we invest in is, I think a -- establishes a standard that is not seen elsewhere in the development community. OK.

We have 10 under contract, 11 more evaluations that are rigorous but without randomization and then 26 others that are currently under consideration. The point being that while we recognize the gold standard of randomization, we also recognize that it's not always possible. What we try to and is maintain the notion that there's a hierarchy of rigor and that we want to use as rigorous a method as possible, and we want to do it as often as possible.

The last thing I want to talk about is our Web site dissemination, and I think that it's important to understand that for MCC this notion of good governance is not just for our beneficiary countries, our country partners, but we see that as part of our model in accountability as a U.S. government agency as well. We believe that foreign assistance should reinforce good government practices, both in our countries, country partners and here as well.

We have an impact evaluation group of more than 20 people within MCC -- economists and M&E staff -- who work on building -- are looking for a broader technical exchange. Let me say that many of them are here and are open to talk with people during the coffee break and after the session. You may contact us at this Web site and that we're posting all of our impact evaluation information on our Web site.

Well, what does that look like? Very quickly, I'll walk you through.

This is our home page where we take the opportunity to explain what impact evaluation is and how it fits into the MCC model.

The next page from that is a sector overview, which allows people to look at each of the different main sectors where we do impact evaluations. This is important because we believe a lot of technical people coming to us will be interested in specific sector issues rather than specific countries, although there will be links to countries as well.

So then once one goes into a sector, then we have a sector strategy page which explains our approach to a factor and our assessment of when to do an impact evaluation, because, let's be clear, impact evaluations are difficult, they're expensive, they can't be done all the time. MCC has to prioritize when we're going to do it and how, and this is described in our strategy page by sector.

Then within the sector page, you'll see the individual evaluations. And here you'll find country tabs and the actual evaluations.

And then my last page will give you an example of what a country page will look like. It provides the logic, it provides the timeline, and, remember, MCC -- I said it's not a new agency, but in a lot of ways MCC is still new. Many of the things we're doing are still in its early stages. MCC is committing to putting our documents on the Web site to provide a basis for following the progress of impact evaluations. These pages describe impact evaluations that are either only early underway or still in the planning stages. We will keep these pages up and update them as progress takes place.

Let me just close with a quote from the Center for Global Development. This gets again to the notion of an impact evaluation gap. "MCC is actively seeking to help fill this gap." We're not alone, and we know many of our development partners are working in the same direction.

MCC, I think, is unique in that this was part of its original model. It's in our DNA. There's a strong commitment to this. The transparency initiative, I think, is critical. It lists our intentions to do impact evaluations, it tracks our performance against our intentions, and it reports our results, whether they're positive or negative. MCC makes a commitment to doing this, and I think in doing that raises the bar in terms of understanding aid effectiveness.

Thank you.

(APPLAUSE)

BARON: Thank you, Franck, for that valuable overview of what MCC is doing and how MCC operates and its work on impact evaluations. Everyone seems to be quoting the Center for Global Development, and we are fortunate to have here the vice president for Programs and Operations and senior fellow from the Center for Global Development, Ruth Levine, who is also a co-author of the Evaluation Gap Working Group report.

(APPLAUSE)

LEVINE: Thank you very much. It's a real pleasure to have the opportunity to speak and to follow Franck after his really helpful introduction to the MCC, an articulation of, as you said, how impact evaluation is embodied within the DNA of that relatively new approach to development assistance.

You know, it's always something when you start talking about evaluation to focus on all the things that we don't know and try to motivate learning more, and instead I just wanted to highlight how much we actually do know.

So there's a lot that development practitioners and policy makers know. So, for example, we know a lot about the magnitude of problems and trends over time in, say, for example, infant mortality, in child health, in the need for family planning services within the health sector, we know a lot about enrollment and, to some extent, completion and dropout in education. So we know a lot about, sort of, the magnitude of the problem and trends over time.

We also know a fair amount about correlates of problems to be solved. So, for example, the relationship of income or rural or urban residence or educational background to key development outcomes in health, education, housing, jobs and so forth. So we know a lot about that.

We know a fair amount about coverage of programs. So, you know, how many schools are out there per, I don't know, 10,000 kids in particular settings and how has that changed over time. So we know quite a bit about coverage of programs.

And we also know quite a bit about operational effectiveness and efficiency, sort of, how well particular kinds of programs are run, whether they're donor-funded, whether they're privately funded through NGOs and so forth. Not everything we might want to know, but we know quite a bit.

But we don't know those things by accident. We know those things because investments have been made over time in surveys -- I'm actually not using the PowerPoint, so that's just the title. So don't wait for the PowerPoint.

So why do we know these things? We know these things because investments have been made over time in large, repeated surveys and in the analysis of those surveys and also in collecting operational statistics about, say, when schools are constructed, when teachers are trained and so forth.

So we know those things because we've actually spent money to learn those things, and they have proven to be very useful.

But then there's what we mostly don't know, and what we mostly don't know is what the direct impacts are that can be attributed to particular programs or strategies, particular approaches. For example, if you're interested in ensuring that girls enter and stay in secondary school, we don't know, with a few important exceptions, we don't really know how well scholarship programs work versus gender sensitization of teachers versus changing the physical plant of schools to make them more accommodating to adolescent girls.

We also don't know the indirect or secondary impact. If we don't know the primary impact of programs, we definitely don't tend to know the secondary impact of programs. So, for example, in microfinance, we tend not to know what the changes are in the household dynamics when women have access to credit, again, with some important exceptions. But we don't know, does that actually make their lives better off in the household of worse off? And what does it do to the dynamics in the community?

And we rarely know whether results that are achieved in one setting, at one time are replicable and enduring. Can they be translated to other settings? Will they endure overtime?

So those are key things that we don't know, and it should trouble most of us who care about how money is spent and how people's lives are affected that we don't know these things.

When we did our work that was published in 2006, we were looking, as people do, backwards, and so we were looking, more or less, around the period of 2004-2005. So this information is a bit outdated and, fortunately, has changed significantly, I think.

But at that time, if you looked across, for example, 600 programs that were funded by the -- financed by the Inter-American Development Bank, only about one-sixth of those collected in their evaluation any information about beneficiaries of the program, and a very, very small percentage of those collected information also about those who were similar to but did not benefit from the programs that had been financed.

A very comprehensive review meta-analysis of payment mechanisms and health care programs in developing countries came up with precisely -- that spanned about 15 years came up with precisely four studies that lived up to standard of validity and reliability to draw conclusions with confidence.

About 10 percent of hundreds of evaluations that UNICEF did attempted to measure impact. And when we did our work on looking at large-scale global health successes, we got 60 submissions where experts in the field claimed really significant success in large-scale global health programs. Of those there were only about 20 that had information that permitted attribution of an impact to the program itself.

And I just want to call out that among the disappointments for that project was that the AIDS prevention programs in Uganda at that time, which were in fact the inspiration for the then $15 billion PEPFAR Program had not been evaluated in a way that permitted a confident -- an ability to draw with confidence that attribution between the approach that was used to HIV prevention and the results.

So who needs to know what works? Well, we'd all like to, sort of, know more. The academics say, "We just, sort of, want to know more," but, practically, there are decision makers who need to know more about what works. Those in the development policy community, both in developing countries and in those who are providing development assistance, to determine the level and allocation of resources.

Perhaps even more importantly our program designers and implementers who are really committed to particular goals -- you know, keeping girls in school, keeping babies alive, making sure there are job opportunities for people who are just at the poverty threshold -- those people need to know what works and how to modify the programs that they design and implement.

And then there are the taxpayers and those who are the intended beneficiaries of the program. So for accountability and, perhaps more importantly, for learning purposes there's a need to know what works. And I would argue, and have at various times that development assistance is always just going to be a drop in the bucket in terms of the amount of resources it brings to the table to solve large social problems in low-income countries.

And so there's a special role that aid can play in providing the resources that can test new ideas in a way that can inform the public policymaking in those countries with their own resources.

So there are a lot of people who would benefit from knowing more about what works, and yet we observe that there are relatively few resources devoted to generating that knowledge. So we looked for explanations of why that was the case and articulated a set of incentive problems that result in underinvestment in impact evaluation. So I'll just mention three of them.

First is that, at least in part, the knowledge that's generated from impact evaluation is a public good. So if somebody does an evaluation of, again, a girl scholarship program in Bangladesh, that absolutely helps the Bangladeshi government figure out how to devote is -- how to allocate its education resources. It also helps education policymakers in Nepal, in India and further afield in Malawi and Tanzania. It has spillover benefits to those who are faced with similar problems, not to, sort of, duplicate precisely but to inform their decisions. So we have, sort of, partial public good being generated.

Second incentive problem is that in any development institution, maybe with the exception of the MCC now, there's some tension between the doing function and the learning function, and I think those who are making decisions and the task team leaders at the World Bank, the program officers and the CTOs at USAID all face this on a daily basis. When they're designing a program, they want to be the beneficiaries of all the knowledge that's been accrued to that point so they can create the best program. But at that moment, they have very few incentives to slow down, to devote the money and the time to designing the baseline for the eventual evaluation that will generate knowledge for their successors. So there's a real dynamic inconsistency problem and a tension between the professional incentives for doing, which are very strong, and those for learning, which are much, much weaker.

And then the third incentive problem is that there are really quite palpable disincentives for candor. Knowing what works constrains political processes, particularly if that knowledge is out there in the literature. It constrains the discretion that managers and politicians can use to direct resources where they wish them to be.

And there's a kind of bureaucratic logic that prevails in many development institutions, making it very challenging to communicate bad news that something didn't work up the chain of command.

There are currently a number of efforts being made, I would say, certainly a real increase in the past two or three years to address these incentive problems, some through very pioneering efforts of individual institutions and some through some collective efforts. I'm just going to name five of them, and that will be the end of my remarks.

So one approach that has been taken is to mandate through legislation that more impact evaluation be done. Mexico is clearly at the forefront of this, but we also see it in our own country. So if you read carefully the language around the reauthorization of the president's emergency plan for AIDS relief, you will find quite a strong push for better evaluation of all sorts, including impact evaluation -- a market improvement in my mind from the original PEPFAR legislation. So you can pass laws, which also involved allocating public money.

Second thing that's being done is that there's a move afoot to have a prospective registry of impact evaluation, must as you do for the MCC-sponsored evaluations but more broadly, an idea that's, sort of, borrowed from the medical world as a way to avoid the potential suppression of results when things don't turn out as those at the top might wish in terms of the results.

Third approach that's been taken is to demonstrate real excellence in the conduct of evaluation and to disseminate that widely. And I'm glad that Rachel Glennerster is here on the second panel to talk a bit, perhaps, about the work that JPAL is doing. And there are certainly others who are really at the front here of methodologies and their application for impact evaluation.

Fourth of five is that there's the International Initiative for Impact Evaluation, which is a new enterprise that pulls together financial resources from foundations, NGOs, developing country governments bilateral and multilateral agencies to commission evaluations around a shared agenda of what are called enduring questions, our evaluation topics.

And the fifth is that there are clearly institutions that are placing increasing value on devoting their technical and financial and political resources to this effort. So we have seen a lot of activity at the World Bank, partially with support from the Spanish government, at NGOs like the International Rescue Committee, really pioneering efforts to evaluate their own programs and to be transparent about their results and to define themselves as a learning organization.

And, certainly, as we heard earlier, the Millennium Challenge Corporation is, within this country, really on the leading edge of impact evaluation. And I hope that their own institutional -- the institutional value that they place on impact evaluation will be a bit contagious to the other agencies that provide development assistance from the U.S.

Thanks very much.

(APPLAUSE)

BARON: Thank you, Ruth.

Those are the formal remarks on the panel, and then we've also asked Mark Lopes, who is senior policy advisor to Senator Robert Menendez, who, as I mentioned, chairs the International Development Subcommittee of the Senate Foreign Relations Committee, to make a few informal observations from here or down there or wherever you want on the remarks, and then we'll open it up for a general discussion.

LOPES: Great. Can everyone hear me all right? Better? Better, better? I'll stand back since sometimes I get animated.

Well, listen, I appreciate the opportunity to come and speak, and I think, just looking around the room, I mean the collective wisdom in here around evaluation, monitoring, managing these programs in a variety of agencies is tremendous. So I don't pretend to provide any additional insight beyond what's here, and the rigor with which you all have spent tackling this problem over many years is both impressive and daunting in a way that we who work in a political context have to try to capture as much of that as possible and try to build it into our programs.

Just a couple of comments, I promise I won't take a few minutes. Maybe on the context which I work now with respect to the MCC model, which is, I think, largely considered, sort of, up for questions -- if that starts to really annoy anyone, just let me know.

I think MCC -- and I'll talk just a little bit about MCC, because I want to, sort of, broaden it out -- the model, from my perspective, was built to try to dodge some bullets around the, sort of, short-term, go in, go out, five-year mandate. It tried to distance itself from, sort of, the congressional winds, meaning earmarks and, sort of, a year-to-year interest on the part of Congress to -- maybe it's easier if I just bring this closer.

And also to try to limit their scope in terms of they're only going to work in a few countries. They are going to build on governments that are already doing well, basically, and they're going to build on -- build host country capacity.

Now, I think in all of those areas MCC has evolved, to some degree, to -- I don't mean find out -- but to need to take a step back with respect to all those areas. I mean, the short-term issue with respect to countries want a concurrent compact or subsequent compacts realizing that five years isn't enough for a lot of the work that needs to be done. With congressional involvement around, some of the political factors that I think are coming into play I think it's important that MCC stay with the model that they have been built on, because that's the biggest justification for a lot of the ways that the program is designed.

The scope of the program, I think the Threshold Program, there's a few other examples that I can talk about where MCC is finding, for very real reasons, that it's nowhere near as neat and analytical as had been hoped. And then in terms of building on the country capacity, I mean, I think, no one's going to disagree with that's the ideal way to do things, but we've also found that project implementation units and a time factor that they have to get stuff done is very real, and there's some sort of smart balance in between the two of those whereby you build on capacity as much as possible but then also make sure that you're not sitting around for years.

So I think they've not been able to, sort of, avoid those hard questions, but on the positive side or, sort of, on the positive side I think MCC has raised the bar around the analytics and the impact evaluation in a way that I have not seen in terms of the buildup before compact and then trying to look at things across a compact. They've raised the bar in a way that's tremendous and fantastic and I think it's going to be a valuable investment for the entire development community, and so it's important that we continue with the -- and follow through with it -- to see what's happened.

Now, along with that, I think, since they've raised the bar, they've been criticized for not having achieved it right away, and I think that's expected. You know, unfair is hard to say because that's the bar that they set for themselves. But I think the fact that they have -- so when you put the bar here and they've gotten here where everyone else has the bar here and they're above it, that's not appreciated very much. And so I think there needs to be some sense of like, "Listen, we made this game much harder for ourselves, and we're trying to advance this farther." So I think there needs to be a little bit of acknowledge there.

I think the F process within AID, and State also, provided some very useful information to Congress, and they've gotten nothing but criticism. There's a couple of things that we can now get in terms of budget numbers across countries that we've never been able to get before and now we get a little bit and we want more. So just to acknowledge that.

And I just want to say that there seems to be a disconnect between the academic community, which we'll hear in the next panel -- or I see the disconnect. A lot of this conversation -- the analytics, the cost-benefit, the real, kind of, think tanky academic world -- it's a world that I don't work in right now. I mean, this -- I deal with these issues every day, and I rarely have these kinds of conversations. And so whether I'm hanging with the wrong gang or not getting out enough or whatever, but nobody that I know closely has these conversations frequently.

So we may have them for 20 minutes, we may have them half an hour or an hour, we may go to lunch, but this is not something that consumes a majority of my day, which I think is an issue. It's a problem and we can go into the details, but I just want to acknowledge that, that I think we're merging worlds in this room, which is fantastic, but they're worlds that operate, from my experience, largely independently. And we come together for lunches and we come together to talk about how important it is that we work together, and then we go off and we work in our separate institutions.

The last is, you know, the political context around putting this in, I just want to give a couple of examples. You know, you're talking about a real big picture, sophisticated level of houses (ph), so I just want to mention one program that I think is interesting that AID just e-mailed me the summary.

It's a study that AID did on democracy assistance programs around the world in 165 countries, from 1990 to 2004, to find what investments in democracy resulted in, like did this have any impact on democratic principles.

And so the way that they set it up was that they had some academics come in a study this, and then they had another team of academics paid for out of the same pot that evaluated the academics who did the study, which I think is a fantastic way to build in objectivity. You know, you have people's reputations on the line, which is, in my mind, the most -- the best way to create an incentive. And so they looked at the methodology that the first group looked at and then they came up with their conclusions.

So what they found was, basically, that democracy funding -- for every $120 million in USAID democracy and governance funding produces approximately a five-fold increase in the amount of democratic change over what the average country would otherwise be expected to achieve.

So I don't know what that means, but...

(LAUGHTER)

...it sounds like it's a good use of money to me.

So I would just, sort of, put out there, one, how many people have heard of this study; two, who cares? And, you know -- or do you care and have you really been working on -- and I say that not to be glib but to say that democracy funding is one of those pops of money where everyone, sort of, looks at it and says, "Well, you know" -- I mean, I think you have strong views from either side that are unsubstantiated assertions, basically, and I'm here to, sort of, offer a few additional unsubstantiated assertions around this work.

What does that mean, and will a study like that, no matter how rigorous it is, actually result in policy changes in the next whatever? So even if we could capture the value of this and put it into a spot that makes perfect sense, how do we built that into the next decision?

And tying into that, until I started working in the capacity that I'm in now, I never realized or appreciated the number of people that come to me who insist, absolutely insist, that their way of doing things is the best way to do it and that that is a valuable investment for the U.S. taxpayer. And they come with numbers, and they have figures, and whether it's microfinance or women's empowerment or getting girls in schools, there's a deep number of people out there that use the data that they have to try to build their case. And I think -- let me just -- and I think that the political power that's behind those initiatives should not be underestimated as well.

So what we're asking for is, how do we build in an agency, an institution, a U.S. government -- and this is what I think about -- a U.S. government, sort of, engine that can contextualize all these factors, understand the rigor, understand the political forces, understand the special interest groups that have a particular stake in this and make that all make sense. And I'll talk about how we're doing that at a later date (inaudible).

BARON: Good, thank you, Mark.

Let me just mention that on the next panel -- one of the things that struck me when you were talking was the gap that you mentioned between the discussion on Capitol Hill or, sort of, in the policy world and the discussion that happens in academic or among researchers and how there's not necessarily a connection.

On the next panel, there is a very good example from domestic social policy, Welfare to Work, where research findings of the kind we're talking about had a major impact on national policy. So that might help shed light on the subject.

But with that, let me open this up for about 20 minutes of discussion here.

Yes?

AUDIENCE (ph): My name is Larry (ph) (inaudible).

You had said we have lots of development programs, and one of the things that sometimes -- I'm thinking right now (inaudible) where you have to have a minister of finance and just the steps they've achieved of the reconstruction days in that kind of setting. If we (inaudible), the most important people in development aid are auditors and accountants, because they create where the money goes. But when you put up there that things in Indonesia where you had the audit and were able to make it more effective -- the eight more effective. So I'm thinking in terms of this evaluation, in terms of impact, what is being done in terms of getting accountants and auditors who evaluate where the money actually goes?

WIEBE: Well, let me just speak on behalf of the MCC model. Again, what I think is important -- this is why I emphasize the whole, kind of, cold chain of results is that it's not just the evaluation at the end, because often you've already spent the money and you're try to figure out the next round. MCC's model is to start off with an assessment of what you're going to do and what the effect is going to be.

That establishes a baseline for all kinds of performance metrics for all kinds of implementation, in terms of whether you're building a road, what kilometers, what the width is, what the depth is, if you're building schools, if you're training vocational -- if you're doing vocational education for adult learners. There's a series of activities that you have to do. Those are all built into the model in the benefit-cost analysis first to say whether you think that the effect is going to be worth is.

That establishes material benchmarks that you measure progress and implementation against. Now, we require countries to report on their progress on those, and those reports allow us to see not just whether the program is going to have the effect at the end, it tells us whether the implementation is on schedule, where bottlenecks are or where things are not working as we thought, as we're going.

And so, again, it's a -- I'm sensitive to Mark's point about everybody has their own numbers. If we had time, we could talk about who this model is materially different, but it exactly puts in the hands of the accountants and the auditors, if you will, a sense of authority over whether things are on track, whether they're costing what you thought they would and whether they're being implemented according to schedule.

That's the middle part of the story, the monitoring and assessment during implementation, that I glossed over but that is a critical part of our model. In fact, the people who do the evaluation, or the team of 20, including not all but most of the folks that we have working on M&E plans in country. Now, again, like I said, we have international procurement processes to make sure that that's done according to best practice principles. It's exactly because we have an explicit statement of what you're trying to get and how you're trying to get it that allows you to do exactly what you're focusing on.

BARON: I'm going to ask the panel to use the wired mikes and the audience to use this mike.

So other questions? Yes?

SCHROEDER (ph): Hi. I'm Mark Schroeder (ph) from HUD, and I have questions for Franck.

As one bureaucrat to another, this transparency thing can really hurt. Sometimes research fails. It can fail, first of all, because your contractor defaults on you, it can fail because the program administrator, one way or another, defaults on you. Even if the administrator doesn't, the program staff can either sabotage it or just never implement because they don't have the same hypothesis that you do. You may have failed to design a good data collection strategy, and it turns out that the way you were planning to follow up and get your data at the end, that didn't work. People may never have enrolled in this program. There may not have been much demand for it to begin with. Or the political leaders at the top may decide that either they aren't into this program at all after -- there may have been an election and they aren't into the program at all or they want everybody to have it, forget your control group.

Which of those types of failure do you think it's worth reporting to the world on?

(LAUGHTER)

WIEBE: Well, as a bureaucrat with my bosses in the audience here, let me approach this question with some sensitivity, and I'll also put my glasses on.

No, you're absolutely right, that any one of those things can happen. In fact, we experienced some of those in many of our activities, as do other aid agencies. Going back to the model, the cold chain of results, one thing of importance is to recognize that that creates a feedback loop, so you're a learning agency. I think one of the things MCC emphasizes on the impact evaluation front is that we don't want to be making mistakes. No agency is going to proudly declare, "We failed." And I think MCC, we'd be unrealistic to expect MCC to say that as well.

MCC is willing to say, "We failed and these are the reasons why, and this is how we will avoid doing it the next time."

Some of the things that you identified, though, were things that should have been known ex-ante, right? And if we didn't incorporate those into our model, into the benefit-cost model, into the country ownership model that is supposed to protect the country moving forward in implementation through the five-year period, regardless of whether there's democratic change or not, if we did that wrong, well, you know, we did it wrong, and we should do it better, and we should learn from that too.

But I said some of the things you should know, there's evidence. If you say, "I'm going to build a road in five years," and you're only halfway done, you ought to learn from that. The next time you should say, "Well, look, if we do it in this country, especially, we can only build half as may roads."

Let me go back to one meta-study I like to refer to, I think it's in the bank, that looked at particular kinds of programs and looked at 30 activities that they had done in that particular investment -- I'm not going to mention which one it was. It was lines of credit.

(LAUGHTER)

Targeted lines of credit. And the 30 had all had impact evaluation, and I believe -- I may be exaggerating, but it's very close -- half of them never got to a finish line. They failed because they simply couldn't do what they said they would do, and they knew that before they got to the finish line.

The other half, none of them demonstrated cost effectiveness. They had failures in repayments, they had failures in all kinds of ways, right? And the conclusion of that is, "You know, next time we should do it this way," right? If you do it 30 times and not one success, I mean, how dare you come up with a model that says, "This time we're going to be successful."

I mean, there has to be a learning loop. If you do the benefit-cost analysis, if you say, "How are we going to do this and what evidence do we have," it creates a new approach. It says, if 30 times we failed, this time let's do a small pilot for two years and if you can show me that you're effective in those first two years, then let's ramp it up. It's very different than saying, "Let's put the whole thing forward."

And MCC has one program where we use that thinking very much on a large scale and it was a land reform in Burkina Faso, and the country definitely wanted it and we had strong support for it, nobody was opposed to it, and yet we couldn't really pull together evidence which said -- I forget exactly what the full amount was, but it was a large program. And we wanted to support it, but we said, "It doesn't meet our standard."

As a result, what we have done is we put in place an impact evaluation that will take place after the second year, actually, into the third year. And, obviously, you're not going to have change, you're not going to have revolutionized land title practices and investment practices, but we all agreed on what evidence would expect to see after two and a half years to from a $20 million program to a $60 million program, something like that. And we all agree that without that, we would have to have a serious conversation as to why you would do the extra program if you didn't have -- if you aren't meeting the targets you had set.

So, again, it creates a way of using information that says, "I'm going to use what I know, and what I don't know I'm going to learn, and I'm going to feed that back in." And all the things that you said will happen. If you document it -- if you remember my, kind of, closing language, "We're going to tell you our intentions, we're going to track it, and we're going to report on it." No model that fails all the time is successful, no matter how much learning and reporting there is, right? We don't expect that, but we do expect, through our transparency initiatives, to put that information forward.

LOPES: Thanks. Just another thing that came to my mind was things that aren't measured that MCC, in this case, as an example, has been able to achieve. From my perspective, only visiting a few countries where MCC works, speaking with very high-level government officials who have a deep knowledge of the substance of the MCC compact and a focus over how to make progress on that compact and questions about how they can be qualified for this or that is something that I've never seen before. And so they've been able to engage mid-level but also very high-level officials in a way that I have never seen before.

And so how do we measure that, what do we care, et cetera, et cetera, like, how's that going to fly in the Senate hearing? All those I think are mushy, but for those who've worked in this business and feel like that's actually important. It's something that we haven't measured, can't measure, hard to justify broadly, but it's something good.

So there are -- I guess this is another way of saying, this is all a bit messy, in terms of what do we do about this and what we want to do.

To the first gentleman's question around the auditors being important, I mean, I think in the spirit of being contrarian, disagree a little bit with the, sort of, those folks later down the stage being really important, but we're putting our emphasis on the people at the very beginning being the most important conduits and actors in trying to build these programs, those who have a sense of the research base, those who can grab the information from the experts that know it when they're designing programs. I think Ruth was very generous to CTOs at AID.

I mean, I've worked on many of those teams and we've got $10, $20 million programs, small stuff, basically, but they're just not justifiable. Like, if you pushed on the questions and you ask them why they're doing it and you try to figure out, like, how does this fit in in the broader context, et cetera, it gets very murky.

And that was the frustration when I worked at AID, and I think -- so if we can build the kinds of officers, and this is what we've been doing on the committee, in AID that personally have a stake and have an experience base that can just do good work and get our there and make things happen in a smart analytical way, objective, more business students, et cetera. Then I think that dramatically increases our chances of success. So that's a very working level, non-quantitative, kind of, operational, middle management reform that we're trying to do from the inside out of AID to build in people that have the networks and the minds, for whatever reason, that can at least give us a better chance at the beginning.

BARON: Let me just add one note to that in terms of building for success and evaluations. From our experience in working with a number of different federal agencies, one common problem is that sometimes agencies and other organizations evaluate ideas that are not well-implemented; meaning, our advice to the agency is make sure that the program you're evaluating is something where all the parts are operating as they should. You've done an implementation study, you know that people are getting trained, they're getting the assistance or whatever. Before you even consider an impact evaluation. There's no sense rigorously evaluating something that isn't being implemented properly. And that often happens. Number one.

Number two, our main suggestion in terms of selecting an evaluator so that the study will go forward effectively, because a lot of rigorous evaluations fail, is to find someone, a team, that has done it successfully before, not just that it has a good design on paper, which is easy to do, but has a good evaluation requires a lot of organizational skills, interpersonal skills and not just research ability. So to find someone with a demonstrated track record to do the evaluation is very important.

There are some questions in the back of the room, I understand. All the way in the back, still eating lunch back there.

AUDIENCE (ph): Thank you, ladies and gentlemen. My name is Macharia Waruingi. I am a doctor of medicine and doctor of health administration. I'm from Kenya and I am very interested in the question of development. This is very important meeting for me, because what we are discussing today is what we refer to as the implementation gap. We have studied this phenomenon quite extensively, as Africans in Kenya. We perceive this as a problem of relevance.

And the problem truly creates a relevance paradox. And relevance paradox occurs when we create solutions to problems without certain knowledge of what is going on. Dr. Levin was very good at saying that we do not know all of the things happening. And the question then becomes, how do we generate relevant knowledge? How do we know?

Now, evaluation, then, is not the problem as such. The problem is how to know the relevant criteria for evaluation. How do we know that what the criteria that we use for evaluation are going to generate relevant results?

OK, so, now what do we do to avoid the relevance paradox, which is a major tragedy that is widespread across development initiatives. One good thing that Dr. Levine said was that we're very quick at doing, but we're not so quick at learning. So as you can see, this is a learning failure. Doing without learning is also known as type I learning disability, when we cannot learn about what we're doing.

So how do we create a possibility for learning? And creating possibilities for learning is a problem of inclusion. Now, knowledge emerges from people, but knowledge is -- there's nothing new that we can create, knowledge is never created. But knowledge only emerges.

Now, in order for knowledge to emerge, true knowledge to emerge, I am going now to give you the criteria for evaluation for effectiveness. It has to plan from a platform of inclusiveness. Now, what we see in development community is a big possibility for learning is impossible because we don't have the sufficient number of people who are called stakeholders of development included in our platform.

People missing include the business community. They are never mentioned. The international business community and the local business people they are out of deep awareness of what is going on, because in market creation, creating market for your product, you truly have to know your people well. That's one stakeholder of government we totally ignored, and that includes multinational corporations and the local community.

Second, if the local people -- and this is very, very important -- in malaria-controlled programs, they are totally, totally destructive in very many areas. So how do you know what local people know in order for you to create solutions that are relevant to them?

BARON: Does anyone want to comment?

That was a very valuable comment, more than a question, I think.

Let's see, are there are any other questions? Yes?

AUDIENCE (ph): I think he made a very important decision.

BARON: Yes.

AUDIENCE (ph): (inaudible) elaborated on moving (inaudible) forward (inaudible).

BARON: OK.

Anyone in the audience want to comment on that specific question?

AUDIENCE (ph): Just a comment, and this is putting in a little bit of stuff for MCC. MCC does (inaudible) local stakeholders when they're setting up their programs of one of the studies we're evaluating. They try to make sure that all of the local -- stakeholders at all levels, including public, including farmers and mayors and those kinds of folks are involved and kind of have a say in helping inform what the questions are. So I do think that in -- maybe it's not true everywhere, in general, but I do think, to some extent, there is an interest in engagement of local people because they know more what's going on and what the needs are.

BARON: Let me just mention one other thing on that point about involving local stakeholders. Just to -- there was an evaluation, a study that was done by senior World Bank economists a few years ago, which looked at the question -- sort of surveyed the evidence on whether the trend toward greater community involvement in the design of projects by involving local stakeholders, whether there was evidence to support that or against that one way or the other, and that was one of the earlier World Bank studies that I was referring to, which essentially found that that is, sort of, a valuable theory about what works that's out there, a certain strategy, but it's also one where there's very little evidence about whether it actually works or makes a sizable difference or not.

Yes?

MICHAEL (ph): OK. Thank you. I'm Jim Michael (ph). I'm a consultant.

I had a question that touches on some of these same concerns about stakeholder roles, and that is we are now looking at a pretty broad international consensus in favor of an effective aid development effectiveness, managing for results, sort of, agenda coming out of Monterrey and the Paris Declaration and now the Craw agenda (ph) for action.

And I wondered to what extent MCC and others in the U.S. government are taking this dialogue about evaluation and measuring results and measuring effectiveness into the broader international community, taking into account Ruth Levine's point that most of the resources for development do not come from development assistance, and I could add that most of the development assistance does not come from the United States.

Is this being integrated into a broader international effort to make development more demonstrably effective and development assistance more demonstrably effective so that development will be considered as a legitimate issue in international relations rather than a side bar issue while the main event is more in the power politics or the military dimension of international relations?

WIEBE: Let me take the first crack from the MCC perspective. I hope that from my presentation, first is -- the first thing is to recognize that this focus on results is firmly placed within the context of the Paris Declaration on aid effectiveness. And when I talked about the MCC principles, many of those overlap exactly. I think that there are still some open questions, but MCC and the U.S. government is engaged in the Paris Declaration and involved in those conversations. That's the first thing.

The second is that if you read the paper and from my presentation -- you know, one of my objectives from this is to, in some sense, place a new light on that Paris Declaration, say, "You know, even managing for results is really a process statement, and, to my mind, when we're talking about aid effectiveness, we ought to be talking about what are our results, not are we managing for results, but what results are you trying to accomplish and is it worth it?

And so I think the title of my paper is, "Placing Results at the Forefront." Maybe I would -- another way is, "Placing at the Center," "Placing at the Top." In other words, process is important but you can have good process and no results. MCC is on board with the process but we also take it a step further in terms of the results.

Now, on your second question, we're not alone, and this is part of that discussion. My former colleague, Andrew Warner is here. He's now at the World Bank with the Impact Evaluation Group. He also is the secretary general, or something like that, of NONIE, which is the Network on Networks of Impact Evaluation. These are donor organizations that are getting together to talk about how to do more evaluations and better evaluations. That's the core mission of that group.

At MCC, I had the opportunity to be at a recent meeting that was associated with the European evaluators site or something like that. So that's part of it. So it is on the agenda, 3IE is another initiative.

So MCC, I think, is -- you know, like I said, it's an unusual agency within the U.S. in the sense at how central this is to our way of doing business. We're part of a broader community that's going in this direction, and we're part of that conversation that we hope will facilitate an improvement within the broader development community as well.

BARON: And this will be the last question.

LEVINE: Just make a quick comment on this. The discussion about, sort of, harmonization and alignment of development assistance is fundamentally, in my view, a kind of inside baseball conversation that intrigues and delights the people who are really engaged and, sort of, doling out development assistance, organizing their bureaucracies and having these, kind of, global-level conversations. When they go to their parliaments, the parliamentarians and the people they represent are asking, "What difference did you make today in people's lives." And they care much less, I think, about how well harmonized the whole team is and much more about whether there's a significant impact of all the money that's been mobilized.

And in this, sort of, contractionary period that we're in, I think that's going to become an even more important part of the argument to sustain relatively high levels of development assistance that you see in Europe.

BARON: Mark, you want to strike the last word, which I believe is how the members of Congress do.

LOPES: Well, that's daunting but I appreciate that.

On your point, I think there's some -- you know, I wholeheartedly support this push for rigor and push for results, but I think we, as a community, have to be ready for the consequences if we're really serious about doing that.

And just a couple of examples. On the community involvement, if we can't find research that says, engaging a broader audience actually reduces additional results, does that mean we're ready to cut it or is it gone? Are we ready to, sort of, cut that cost out of the contract? That's one point.

The second is, agencies -- AID, government, it doesn't matter, or a bank -- I know of a couple in the World Bank and I'll try to not name any names -- but if you find a group of people, 50 very highly paid, very highly educated people within a large institution that are working on some sort of process statement, if you can't trace the contribution of their work to some result on the ground in the development world, does that mean we cut out that entire office? I mean, I think there are a lot of offices with whom we work that require the justification for the existence that's dubious.

The same goes for this idea of using development and reform as an argument in the context of, sort of, terrorism or broader security. I think there's some implications with that as well, and those who support additional development assistance want to believe that reduced poverty increases stability, makes us a safer country, et cetera, et cetera. If we go back and look at the numbers, are we ready to, sort of, survive the consequences of that? If it says, "Well, actually, it doesn't," are we ready to take the cut?

So I think my only point is that there's politics involved, and sometimes politics can save us from a lack of numbers. And there are some pressure that are in place, and it's part of that whole conversation. So I just want to caution on how far we're willing to take the real rigor.

BARON: OK. Thank you.

Thank you to our panel. If you could join me in appreciating...

(APPLAUSE)

We're going to take about a 10-minute break and come back with the second panel, which is going to discuss some concrete examples. Can evaluations build valid knowledge about what works? That's of policy importance.

(BREAK)

BARON: OK. We're going to go ahead and get started with the second panel, if people could come in and take their seats.

OK. If everybody could come in and please take their seats, we'll go ahead and get started.

If folks from the back room could please come on forward, it would be very much appreciated by all of us up here.

OK. We're going to go ahead and get started with the second panel. This panel, as I mentioned, is going to speak to whether evaluations can, indeed, build knowledge that is usable and relevant for policymakers and scientifically valid examples from development policy as well as other areas, most notably domestic welfare policy.

Our first speaker is Rachel Glennerster, who is the executive director of the Abdul Latif Jameel Poverty Action Lab at MIT, also called JPAL. Also on the panel is Ron Haskins, who is senior fellow at the Brookings Institution and was the lead staffer for the U.S. House of Representatives in the 1996 major welfare reform legislation.
And our third speaker is Dan Levy, lecturer on public policy and faculty chair of the Master and Public Affairs Administration -- Master and Public Affairs Program at the Harvard University Kennedy School of Government.

Rachel, you're first. I'll get your slide presentation up.

GLENNERSTER: OK. Great.

So just one comment I had on the discussions earlier today, I always think a good example of where non-randomized evaluations have served us very poorly in the development area is microfinance. I think it's a good example of why you need a rigorous impact evaluation. Most of the evidence about microfinance in the literature comes from comparing in a given village the women who took up the opportunity to have microfinance and the women who didn't.

Now, if you think -- you don't have to think very long to realize there's probably something very different about the women who, when given the opportunity to have microfinance, will come forward and say, "Yes, I want to learn," and the women who, in the same village, who had the opportunity, decide not to come forward and get that loan.

You know, you can control for income, you can control for education, but if you think about what microfinance is looking at, the ability to start a business, to make that business effective, what are the characteristics that are going to make a woman go forward and take up microfinance? It's persistence, it's ambition, it's entrepreneurial activity, exactly the things that may determine an awful lot else in her life. And so comparing her to a woman who looks the same but doesn't have motivation, ambition to go and get that loan is a very incomplete way of assessing the effectiveness of a program.

So I want to talk about the work that we do at JPAL, and I was asked to talk specifically about some examples of how research has translated into action, which is our motto at JPAL.

So just to tell you -- for those people who don't know about JPAL, it's a center at MIT (inaudible) department, but it now has affiliated researchers throughout the world who do randomized impact evaluations. So, Dan Levy, sitting on the panel here, is one of our affiliated researchers even though he's not at MIT.

Our goal -- and I was thinking earlier today -- is, actually, very similar to the goal of the Coalition for Evidence-Based Policy, which is fighting poverty by ensuring that policy decisions are based on scientific evidence. So we do that by running randomized impact evaluations of poverty programs. Not that we think the randomized evaluation is the only way to go but this is our particular specialty, and we think there's room to do more of it.

We build that capacity of others to do randomized evaluations, particularly in the area of development, and we disseminate the results. And we currently have, if you take the researchers in the network, 70 ongoing projects in 22 countries.

So if you think about the topic of this panel, I think behind it, which is, can research translate into policy movement on the ground, I think behind that question are, sort of, four important questions or implicit, sort of, questions in people's minds, which is, can we rigorously test enough programs to build up the evidence base that we need to make policy more effective; can we answer the questions that matter or can we only answer certain questions and not enough of the questions that are really important; how do you move from one study to policy decisions; how do we make policy going forward, because I think all of us are in this not just for, kind of, evaluation to see what we did afterward, but as people have talked about, learning organizations, about how do we make our work more effective going forward; and, finally, are we just producing this evidence and it's going into a black hole, are people actually going to respond to the evidence when it comes up?

So this is, you know, our propaganda about what we're doing, but I think it also helps answer that first question, the first two questions, actually: Can we do this kind of randomized impact evaluations on a scale in covering the areas that matter?

So these are just ongoing randomized impact evaluations being done by people in the JPAL network. And you can see it covers a lot of the world. It also -- as I say, we have 70 ongoing. It also covers all sorts of different questions. We could categorize these in other ways. I realize we didn't have corruption up there, but that's under participation in corruption, some of the participation questions are corruption.

But education and health are increasingly seeing that this is open to rigorous impact evaluation, but so is rural development, so is participant decision-making, so is gender and discrimination.

So just a couple of examples of projects I'm working on at the moment, what is the best way to empower adolescent girls in rural Bangladesh. We're working with Save the Children, comparing a whole range of alternative ways to empower girls and seeing which is the most effective.

And also we're looking at a project to build trust in a post-conflict environmental in Syria and building trust and building participant decision-making, can you bring youth and women into the decision-making processes in villages, given that these were some of the reasons, particularly the lack of a youth voice, which was seen as one of the key reasons for the civil war in Sierra Leone?

So I want to take a couple of examples and talk about where we've had policy impact and what I think are some of the reasons why you have policy impact. One of them you saw in your handout was an example of a project that was very effective, and this is de-worming.

Now, first of all, it was a question that matters, right? Can we ask questions that matter? This is a problem that matters. Four hundred million children worldwide have intestinal worms. A quarter of the population of the world, actually, have intestinal works, so this is a huge question.

And there was a rigorous impact evaluation in Kenya that looked at the impact of providing school-based mass de-worming. So it's very cheap to do, diagnosis cost more than treatment, so you just test whether an area has worms and then you mass treat. And we saw 25 percent reduction in school absenteeism, long-term school outcomes, so there wasn't initially an effect on test schools, but as these children are being tracked over time, you see long-term differences. They're going to better secondary schools, in particular, is what's coming out.

It's extremely cheap. It costs $0.50 cents per child to treat children for de-worming.

I think the other key thing that is the reason why de-worming has been picked up and is now being expanded is that it wasn't just in Kenya, there was evidence in other countries that this was a very important issue and had similar impacts.

So one of the things that happens to you if you have intestinal worms, is you become anemic and then tired, which explains why kids aren't going to school when they have intestinal worms. And there was a program in India that addressed anemia and also gave de-worming pills, but it was probably the iron pills that had the effect here. So addressing anemia directly also had an impact on school attendance in India.

And, interestingly, in the U.S. South, Hoyt Bleakley, from the University of Chicago, went back and looked at a big de-worming program in the U.S. South where the Rockefeller Foundation wiped out hookworm in the South of the United States. Very similar impact, big educational impact. And because it happened such a long time ago, he was able to trace through the impacts on income from the greater education that people in the South had because as kids they were de-wormed.

Twenty percent of the difference in income between the North and the South of the United States in 1900 was due to hookworm. So just think about the implications for the gap between developed and developing countries now in terms of worm load.

And the other thing that I think was the reason why this study has been -- has an impact is it's a question that matters, we found similar findings in different context, and we can compare alternative ways of trying to get kids in school and look at what is the most cost effective way of addressing that goal, which is, after all, the Millennium development goal. So it's something that the international community has signed up to saying is an extremely important question.

And we looked at all the different randomized impact evaluations on school attendance, and we found that de-wormin was the most cost effective way, cost was $0.50 per additional year of schooling-induced. So maybe we could go to the next slide and I'll show you. Actually, if you go to the next slide, and then we'll go back.

So this is looking at all the different things that people have tested to get universal primary school enrollment. You can provide free school uniforms, you can provide girl scholarships, you can provide school meals. This is actually school meals to preschool children. It would be great if someone did this for primary schools.

And Progresa, which is a very popular conditional cash transfer program, which is also done as a randomized impact evaluation, just to give you a sense of the scales, because we can't put them on one graph because they're so different, this is $6,000 per child per additional year of schooling-induced, and this is $3.50 per additional year of schooling-induced.

I am not saying in this graph that you shouldn't do Progresa. Just don't do it as a way to get kids in school. Progresa conditional cash transfer is about giving money to poor families conditional on getting their kids in school. It's a great way to give money. It's not a very cost effective way to just target getting kids in school. That's the point of that graph.

So if you could just go back to the previous slide.

So they say this is something where people are responding. We've seen a policy response in Zambia, Ethiopia, we're working with a big microfinance organization in India, one of the biggest, SKS, and they're looking at whether they can get de-worming pills out through their networks. Madagascar is starting a big de-worming program. And 10 million children this year will be de-wormed as a result of efforts that we're aware of, which we think are pretty closely related to these results coming out.

OK. I want to give one other example of policies being effective and one of the studies here is by Jessica, who's in the audience. There's a big debate about pricing, sustainability and access around health issues, particularly health prevention, and there's a study that, as I say, Jessica Cohen of Brookings and Pascaline Dupas at Dartmouth, now at UCLA, did, which looks at increasing the price from $0.0 cents to $0.75 cents for providing insecticide-treated malaria bed nets. That led to a 75 percent drop in pay cut.

Maybe you could go to the next slide.

Again, we're seeing very similar results in other areas, so charging for de-worming pills, very cheap pills, but the whole program collapsed when they charged for them and the act of paying for chlorine. So the argument that you should charge is that people don't value something unless you pay for it. And all of these studies show that this is not -- this has not been shown to be the case on the ground. People didn't use the malaria bed nets more if they paid for them. If you take a pill, you use it, right? So there's no difference between take-up and use in those pills, and paying for chlorine to chlorinate your water did not lead to greater use in the study in Zambia.

So the take away from all of these different studies is that small payments that people have advocated in the development community to build ownership, to build youth, to build the fact that they want these products is deterring use, it's deterring take-up and not increasing use. And, again, we have a policy response. PSI, who's one of the big providers of prevention programs around the world, stopped charging for bed nets in Kenya and Somalia in response to this study.

BARON: OK. Thank you, Rachel...

(APPLAUSE)

... for those excellent, concrete examples of how these kinds of evaluations have produced knowledge that's policy importance.

Our next speaker is Ron Haskins. Let me just mention that he's worked in an area, Welfare to Work, which is really the one area of domestic social policy where there's a large body of evidence in these kinds of studies that had a major impact in policy, and he's going to talk about that, I hope.

HASKINS: OK. The bad news first: Random assignment shows that programs often don't work. So I'm tempted to say, programs usually don't work. The great American evaluator, now late great evaluator, Peter Ross at Amherst, had what he called the iron law of evaluation, and the iron law of evaluation is that for any given social intervention, the expected impact is zero.

And in my career, I have learned that lesson over and over and over again. In fact, as you can tell by my hair, I'm approaching retirement, and when I'm retire, I'm going to organize program evaluators, and our motto will be, "Stop random assignment evaluation," because it's too dangerous. It might show a program doesn't work. If you don't have evaluation, you can be like a member of Congress and claim this program will do everything, include slice bread, so let's not evaluate, it's too perfect to evaluate.

But for those of us in the real world, there are, I think, actual examples -- and this is why Jon asked me to come today -- of evaluations that have actually affected policymakers.

Now, this is kind of a smart aleck model or this is something Jon is doing to obstruct my presentation.

(LAUGHTER)

So try to ignore that.

And, normally, when I -- this is various factors. This is actually a model that I just made up late one night, and it's all kinds of factors. And for anybody here who works on the Hill, you know that there are all these -- every one of these factors can be really important. Of course, a chairman's here, so chairman. The chairman can definitely be the most important, can dominate the whole thing and often does, especially in previous Congresses. Before the leadership of the Democratic party and Republican party got so powerful they often did.

And here is research. I normally do this is a little slice here, but in deference to Jon, I expanded it here.

(LAUGHTER)

So this is no longer based on my great computer model, it's just kind of made up. And, of course, it varies from time to time. But I like to think of my mission, and the mission of many people at Brookings too, to expand this piece of the pie. Because, at least potentially -- this is really hard for a lot of people to agree to -- but at least potentially, social scientists are the least biased people in a policy debate.

Now, I have -- I was with the Ways and Means Committee for 14 years, and I listened to many highly reputable social scientists -- if I named their names, many of you would know them -- who came to the committee and said all kinds of outrageous things, because they're advocates too, and maybe all of us have a little bit of advocacy in our heart, except I'm a Republican, so I have no interest in children and I can just call them as they are. But everybody else does have bias.

But this is a very serious point, and it goes to places like the masters degree program at Kennedy and we have like a 100 hundred of them around the country now that masters degree students should learn, that evidence should play a big role in policy, and it won't play the right role unless the people who interpret the research for policymakers are honest and go with the numbers, so to say.

So let me give you this example of welfare reform. It began -- and I still have not anybody who knows how this happened, but in the Social Security Act, in section 115, some green genius put in a section that said, "The secretary of HHS can waive all kinds of laws if it will increase knowledge." Isn't that great? Would you think that the federal government -- in the early '60s.

So this led, eventually, after a couple of decades, to lots and lots and lots of waiver experiments, and we still have it. We have them in child protection programs, we have them in Medicaid, we used to have them in welfare, you don't need them in welfare anymore because the states completely control and the statute basically says, "States do whatever you want to as long as you meet the work requirement."

But leading up to welfare reform this was really critical. By the time welfare passed in '96, 41 states had evaluations, many of them random assignment evaluations. And, indeed, I think this has played a huge role in the culture of evaluation in the United States, because it stimulated the creation and expansion of great research companies like MDRC, Mathematica, APT (ph) and all of them who were involved in really very terrific random assignment evaluations, often multi-site, thousands and thousands of subjects. You can review this literature, and you can have ends apply, 50,000 across all the studies, because that's how many were in randomly assigned experiment control group.

And let me just say -- next slide -- that from this literature -- and I couldn't summarize it completely -- but this is not an atypical example. This is a random--assigned experiment done at three sites across the country, experimental group, control group and this is just impacts on income after two years. And as you can see, very substantial impacts, random assignment, multiple sites, very good data. So there really was a big impact.

And the intervention really is quite simple. It's not training and education and so forth, it's telling them they've got to get a job and then saying, "We're going to help you find a job. And then when you get a job, we're going to help you find child care and things like that." So quite a simple intervention, and yet it produces big impacts. Next slide.

I want to say a little about the welfare reform bill now, but before doing that, I want to tell you about a provision we put in the welfare reform bill that Jon has copied on many occasions, and I think he takes it to be about the second most important goal in his life to try to get language like this in every single piece of legislation that passes, including from the foreign affairs appropriator. I once sat in a room listening to him regale the poor staff members who were doing the foreign affairs bill, and he persuaded them to put in language on random assignment for World Bank projects and fund them. So I don't know if they've actually done it, but I was pretty impressed by that.

So the secretary shall to the maximum extent use random assignments as a way of methodology. It's in the statute. We also tied it to $25 million and gave the secretary almost complete freedom in how to use the money, and he's doing some (inaudible) and other things, but a majority of this money is being spent to do random assignment experiments on important questions raised about social policy by social security at (inaudible). Next slide.

Now, the welfare reform bill, based, in part -- this is politics, after all, think back to the first pie chart I showed you -- but this is based, to a substantial degree -- on the idea that people should work, and we now have good evidence that if we help them find a job and do fairly simple things, they will actually work.

And, in fact, interviews began to show they want to work, even though they've been on welfare 10 years or whatever. And people would stay on welfare for a very long time, but if you arrange administration so that you tell them they should work, you're going to help them work, you're going to help them find a job, they will actually do it.

And so we came up with this bill. We ended the cash entitlement. Some people might call that a negative reinforcer that we said, "You can't stay on welfare forever, because you don't have an entitlement to it." No one deserves welfare, you have to work to get it. It's a bi-directional effect between you and the state. We're going to give the states block grant.

In the old days, if they got people off welfare, we helped them by cutting their federal dollars. If they brought more people on welfare, we gave them more money. Somehow those incentives didn't seem exactly right. So we gave them a fixed amount of money. If they could help people get off welfare, they kept the change, so to speak.

Strong work requirements, mandatory on the states, and they would be sanctioned. They would lose part of their money if they didn't meet the work requirements. And the same with the individuals. If they didn't do what the state told them to do, then their grant would be cut, and eventually it would be completely terminated, and there was a five-year limit, which turned out to be symbolic. Next slide.

Then what happened was not just because of welfare reforms, a good economy, many of you may recall, and there were other provisions in law that actually made work a very good deal. Most low-wage workers -- mothers that left welfare -- got a job for about 8 bucks an hour. But then they got the EITC, they had child care, kids got Medicaid, they got food stamps. We made a lot of changes in food stamps to make it easier for them to get it. So it really was a good deal to go to work. Next slide.

And the welfare rolls, which virtually never declined in the past, just fell like a rock. Next slide.

And work went up. These are married mothers, so this is just a trend that started roughly at the end of World War II and gradually picked up speed, and more and more mothers went to work. Here's single mothers, and here's welfare reform. Big increase in single mothers. But here's the most impressive, never married mothers, who were the most disadvantaged, the least likely to have work experience, and yet a 40 percent increase over four years in actually having a job. This is not looking for work, this is they had a job. So this is tremendous success in getting mothers to work. Next slide.

And now if you look at the income of the bottom 40 percent, which includes virtually every mother that used to be on welfare, because they're working for not much money, and as you can see, starting in '93, before welfare reform, huge increases in their total income, which includes earnings and other factors as well, but not so much welfare, because here's the welfare income. They're welfare income is going down, but their total income is going up during this whole period. So it's primarily because of earnings and earned income tax credit, those two things.

And it fell a little bit after -- you're getting tired of hearing from me, so you figure if you do the slides faster, I'll get done.

(LAUGHTER)

It fell a little bit, but notice the difference between this and this. So it's still -- even after the recession, there still is a major impact. So a lot more work.

Next slide shows poverty and here is -- this is poverty and married couple. Families, you would not expect them to be kind of a control group, so to speak, comparison group. Well, look at single mothers, very substantial decline in child poverty, by far the lowest poverty rate ever for female head of families, also the lowest ever black child poverty rate, almost the lowest for Hispanics, and that's quite an achievement. And so every bit comparable to blacks, because so many new Hispanic immigrants who have low skills, and so you would expect their poverty rate to go up. So a big decline in child poverty as well.

And even after the recession in 2001 increase, it's still far below where it was before. Next slide.

So from this, you conclude that those random assignment evaluations that so greatly aided the congressional debate and persuaded almost everybody, including a lot of people, politicians are sometimes like this, you probably don't know this, but they're really stubborn, and they believe things based on their experience and their philosophy and what their mother said and so forth, and trying to shake them out of those beliefs is hard, and I think a lot of people who for a long time thought work -- yes, we like work but, you know, these mothers, they don't have enough education, we've got to get them education first and so forth. It turned out that with all the evidence that I think a lot of them began to change their mind somewhat.

And since welfare reform, there have been a whole host -- there's another whole slide, do it again -- and all these represent random assignment studies that have shown impacts -- or mostly shown impacts, a few did not show impacts, on very important outcomes for children and young adults in the United States.

So I think this is definitely an example of, a, that you can perform these experiments on large groups of people and get interesting results; b, that people will actually pay attention to the results and sometimes even change their mind; and, c, make the right decision.

Thank you.

(APPLAUSE)

BARON: OK. And our final speaker is Dan Levy from the Kennedy School of Government where he teaches program evaluation, among other things.

Dan?

LEVY: I'm a little bit of a bind here, because if cognitive scientists have it right, you will remember by next week about three things you heard today, and I'm by far not the most engaging speaker of today. So I'm going to try to make a couple of points very succinctly here.

The question posed to our panel was, can evaluation build scientifically valid knowledge about what works, studies of policy importance. I think the answer to that question is most definitely, yes, and Rachel and Ron and others in the panel have actually given very convincing examples that that's the case.

But there are also many examples of situations where research and evidence don't affect policy, and so, undoubtedly, part of the reason is the pie chart that Ron just put forth. Program funding decisions depend on things other than just evidence.

And so what I would like to do today is to at least talk about the part that we, as a community, concerned with evaluation and the use of evidence may be able to do to build evidence that is more likely to affect policy. So recognizing that other factors affect policy, but focusing on how we can build evidence that actually does. That is more likely to affect our policy.

And for this I turn to the literature on the, actually, the effect of research and evidence on policy, and there have been many people who have written in this literature. Carol Weiss was certainly influential in the '70s and '80s and other people have written. But I'm just going to focus on what seem to me three crucial characteristics of evidence that would help meet these criteria of being likely to influence policy.

The three criteria are: One, that the evidence needs to be credible; two, that it needs to be timely; and, three, that it needs to be relevant. And so what I'm going to do now is spend a little bit of time on each of these three criteria and then leave enough time for all of you to participate in the discussion.

So in terms of credibility, we talk, essentially, about the scientific quality of the evidence, and part of the idea here is to what extent can we attribute differences we observe between groups to the actual interventions? To what extent can we draw causal inference? And I second what everyone has said before. The use of randomized trials have been pretty influential in this area, but so have other, sort of, good, known, experimental methods.

But I think the evaluation gap report that Ruth Levine and others wrote from the Center for Global Development provided a very rough -- a very thorough, sorry, review of the evidence of to what extent do we actually have credible evidence in a lot of the areas that we care about? And I think that the report basically concluded, well, there are lots of areas where we don't know much.

And what I want to say here is I think we have a long way to go still, but I think since the report was written there has been quite a bit of movement in the right direction. I think there are agencies, such as the World Bank, the Inter-American Development Bank, certainly MCC, the impact evaluation, international impact evaluation initiatives, who have actually launched initiatives that I think have changed completely the way that we, sort of, think about impact evaluations and how we, sort of, fund them.

And so these initiatives, I think, have done something that's very important, which is to not only define what's a good and rigorous impact evaluation, but to design the process of managing evaluations in a way that the evaluator is involved from the very beginning. I think that has been really crucial. I think evaluations in the past -- the model evaluation was, "My project is finished, can you come and evaluate it?" I think that's not the norm anymore. Again, I think we have a long way to go, but, certainly, there has been progress here.

I think MCC deserves a special mention here. I think its efforts to transform the use of evaluation in development assistance have been quite ambitious and commendable, and I don't say this because they're hosting this event, although hosting this event may be a sign consistent with what I'm suggesting.

And so while I think there are many challenges ahead, I think on the credibility, both on the demand for impact evaluations but also on the supply with a number of agencies, such as the Poverty Action Lab, Mathematica, MDRC, APT (ph) and many other agencies who can actually do this work rigorously, I think we've made progress.

So my conclusion on the credibility issue is I think there's lots of -- we have a long way to go but there's lots of energy moving in the right direction.

In terms of timeliness, I think this is an issue that I'm a little bit less optimistic. I think the timelines of the average politician or policy maker needing to make a policy decision on funding and the timeline of these typical evaluations are not always consistent with each other. And so I think there have been successful examples -- Rachel describes some, Ron as well -- but I think this is a challenging area.

And I think in terms of what can be done as a policy community is to, sort of, structure the evaluations in a way, when possible, in a way that at least the short-term impacts can be used to inform policy, because this is sometimes that's all you can know by the time that you need to make a decision.

But even that is difficult to do. We can take the case of MCC, for example. There is a threshold program. That program lasts for two years, and then after that program comes a compact program. And if you really wanted to use the evidence on the evaluation of the threshold program to inform the compact program, you would need to have the results of the evaluation in a time that is unlikely to produce credible findings. And so how to deal with that is a little bit difficult, because there's no clear manner in which the length of the threshold could be modified or anything like that.

But I think my main point here is I think there are ways in structuring the evaluation that could help address at least a short-term impacts that could make the evaluation findings timely.

On the timeliness issue, and this is something related to what Rachel just mentioned, I do want to say that we often think of evaluation in this country is going to inform the policy decision on this program, on this country, and evaluations are probably good. So you would think that even if it didn't affect the funding decision for that particular country, in that particular moment, you would still learn in a way that is important.

And then the last part in which -- or the last characteristic that I think is important for evidence to have in order to be more likely to affect policy is that it needs to be relevant. And I think here others have had several comments with respect to this. The only thing I would like to say is that I think it's useful to structure the evaluations in a way that they answer the question that is in the policy arena, and I think with a lot of the recent impetus to do impact evaluations, sometimes that point gets missed. And so I think the policy questions should be the main driver of what the evaluation question is.

And so here I have a couple of examples, one, I think, reflecting a good, sort of, outcome and the other one a not so good outcome.

The good outcome was an evaluation that Mathematica Policy Research conducted in Jamaica, and this is an evaluation of a conditional cash transfer program, and it was very clear from the beginning that the question of whether this program was going to exist or not was not a policy question of relevance. This program had been funded because of the experience with Progresa and so on, and so it wasn't even in the discussion, "Should we have this program or not?" The more relevant question was, perhaps, "Should we expand this program beyond what we have done?"

And so I think what the evaluation was able to do was to estimate the impact of the program on the marginal participant, precisely on that participant that you would give this program to if the program were expanded. I think that's an example of an evaluation that is trying, at least, to address the policy question that is on the debate.

A maybe not so successful example relates to just a simple phone call that I received about a year and a half ago of someone who had learned all about the techniques of doing impact evaluations and particularly of randomized trials, which, again, I second everything that everyone said here. I think it's a gold standard, and it's the evaluation technique that we should choose first. But it was very clear -- this person was working in a government of a certain country -- it was very clear after 10 minutes of conversation that this person was not so much interested in answering a policy question but in seeing how she could fit a random assignment design to a question.

And I think while this doesn't happen very often, there's a risk of us -- and especially us evaluators have a tendency to try to find the most credible design. But, again, I think as a policy community, focusing the most credible design on a policy question that is of interest is pretty important.

I think I said enough, and I will sit down.

(LAUGHTER)

(APPLAUSE)

BARON: Thank you, Dan.

We're going to open it up for question and answer in just a second. I just want to say one comment, which is going to try to bring together some of the points that were made across different policy areas on this panel. What I'm about to say will either be elegant in synthesizing everything or a disaster. You guys can be the judge, and mark it on your feedback forms.

Incidentally, we do have feedback -- before I make my elegant comment, we do have feedback forms at the end of your packet. It would help us immensely if you could just fill that out. It's just a couple of quick checkmarks and so on and any comments you would have. It would help us in thinking about future forums and improving over time.

The second thing I'd like to do is -- is Leslie McElligott here? Leslie? Yes, if you could just step forward for a second. Leslie on our staff has been the project director for this project. She's the one who has worked with MCC and really brought together all aspects of this forum. So let's give her a round of applause.

(APPLAUSE)

My comment, and then I'll open it up, is this, that -- it's really more of a thought for discussion -- why in certain areas has research had, and continues to have, an enormous impact on policy and have a constituency among members of Congress, Democrats and Republicans, a strong constituency? Like, for example, in medicine, NIH is funded to do medical research at over $10 billion a year, and the only question each year is how much that budget, whether it's going to stay the same or increase over time, and there's broad bipartisan support for that.

I would suggest that one of the -- and there's no question, it's got support from everyone on the Hill. I would suggest that one of the reasons touches on some of the things that were talked about here. Everybody knows that this kind of research produces answers about what works that are really important. They know, for example, that the research funded by NIH, including providing the final confirmation that something works, the large randomized clinical trials that NIH funds, like the women's health initiative, have proven the effectiveness of interventions that have reduced the rate of leukemia -- death rates from leukemia, Hodgkin's disease, breast cancer, heart disease and stroke and had an enormous impact on public health over the past 30, 40 years.

In welfare policy, you had something similar. You had a set of findings that produced clear evidence that moving people quickly into the work force worked as opposed to sending them back for basic education. And I think in development policy, you have the beginning of what is a concrete example and maybe a few examples that Rachel mentioned, which is the de-worming experiment, where people, policymakers, can say, "Yes, there was a study here that was done, I can understand the design of the study and produced a result that is of unquestioned policy importance." I think examples like that, I would suggest, may be a key to building a constituency over the long term, long-term support for this kind of evidence-based approach to many different areas of policy.

And with that, let me open it up to any questions or comments or discussion.

Tom?

AUDIENCE (ph): Yes. Just for myself -- I'm Tom (inaudible), United Federation of Teachers and (inaudible). It seems to me I'm coming away from this conference with three general things that seems to be the issue. One is money, two is politics, three is hierarchy, and I think that was addressed by Franck and Ruth and Mark and also by Ron, to a degree, and Rachel and Dan.

Taking that and taking from what Jon said, it seems that the culture for setting up evidentiary research or randomized control trials in areas that are non-medical or possibly non-welfare reform that the culture isn't there. In other words, we don't find any problem throwing money at polio vaccine or throwing money at measles or small pox vaccines. I mean this is something that's research, is provable and has had dramatic effects. But the culture to do this type of research in areas of our policy and poverty and so forth and so on just doesn't seem to be there.

And I wonder -- you know, it would be nice if we all could come up with a Marshal Plan that really worked. But I wonder if you could, sort, of, address this. I may be rambling here, but it seems to me that -- how do we, kind of, for lack of a better term, start to sell this in the world and to Congress? And Mark is not -- I don't think he's here -- mentioned that conflict between academics and politicians or policymakers, OK?

So I'm, sort of -- this is what's going on in my head at the moment, and I'm wondering if you could address some of these issues.

GLENNERSTER: I think it's worth realizing that we're -- if you compare where development research is in terms of using randomized trials to welfare or (inaudible), I mean we started a lot later. And a lot of the techniques that we now use to take the randomized trial and make it adaptable and relevant and able to implement it in the development context are relatively recent, so we've been learning a lot of techniques about how to make this technique much more flexible, how to adapt it to the needs of NGOs so that you can take on all their constraints and still do a rigorous impact evaluation. Because we're very often not doing, kind of, a straight control treatment. We're using their constraints and randomizing around their constraints.

So this is all relatively recent, if you think about it. And in some sense, is it that surprising that policymakers haven't responded to the evidence, given that the evidence is only just building up. You know, I was looking recently at our -- we've got some programs now, but five years ago we only had a fraction of that, and we had a fraction of the results.

So when people now come to me and say, "OK, tell me what's the most cost effective in a whole range of areas," I say, "We've got 20 trials going on, but we don't have very many results yet."

So maybe it's not so surprising that the culture isn't there, but the culture is changing. I was just looking yesterday, running through for a donor, sort, of who is taking on board this methodology and who is thinking seriously about evidence. I went to the top foundations in the U.S. and realized that we were in discussions with them about how to evaluate their work for about half of them. Six out of the top eight international NGOs are doing randomized evaluations. They're sending people on our course to do randomized evaluations. We're working with many of them. So if you'd asked me that three years ago, it was one maybe, but now six out of eight of the top international NGOs.

So, yes, there's got to be a culture change, but I'm actually amazed at how quickly it's happening, and I think one of the key things we've got to do in terms of policymakers is we've got to build up the evidence. Because did you see how many things there were on welfare that you were responding to? I mean, those decisions were made after you had, you know, 20, 30, 40 randomized evaluations out there, and we just don't have that base yet. Yes, we've got to ramp it up, and then you'll have more of the information.

One thing I wanted to say is, my feedback from policymakers is, "Don't tell me about one study, tell me about the cost effectiveness of a whole range of different approaches," and that's one of the things that we're trying to do. So you have to have a really good evidence base before that. I showed you the one on kids in school. We have to now do that on getting teachers in school. There aren't an awful lot of other areas where we can do that, where we have 10 randomized trials with different approaches and look at their comparative cost effectiveness, but that's something that we're going to be generating, and we'll have a database, a searchable database up in the next six months or so, so that you can look at comparative cost effectiveness of alternative approaches.

HASKINS: I agree with the premise of the question that culture makes a big difference, and a lot of areas of research do not have the culture to support random assignment and weighting and patience and all the things that we've been talking about.

And I also agree with Rachel's premise that we're making progress, and I would cite what I assume you're the most interested in and that's education. I think the Institute of Educational Sciences had a tremendous impact on the field. I remember when I went to graduate school at the University of North Carolina in education. I'll never forget one professor saying, "The vast wasteland of educational research." And, boy, it sure seemed to me that that was correct.

But now random assignment studies or at least big studies that try to deal with the problems that you have if you don't have random assignment -- and we have them on classroom size, we have them on vouchers, we have them on preschool programs, we have them in Reading First and on good teaching, the effects of having good teaching for several consecutive years. So I think we're making great progress in education. The question is whether organizations, including unions, will pay attention to the results of the evaluations.

SROUFE: Thank you. I'm Gerry Sroufe from the American Educational Research Association, and, Ron, I was wondering if you would expand your analysis of the welfare reform plan, which you were intimately involved in? If you were to write a book about how Congress passes a law, what would have been the role of the research that you cited versus the chairman, the committees, the interest groups?

HASKINS: If I were to write that book, it would be called, “Work Over Welfare.” It would have been published in 2006 by the Brookings Press, and you would all be able to buy it on Amazon for about 20 bucks. And it addresses exactly that question, and there’s no question that research was fifth or eighth or tenth.

I would say in the case of welfare reform, don’t forget what really gave it its initiative was Republicans taking over Congress. Republicans took over Congress in ‘94 and, literally, the day we walked through the door in January we introduced that bill because we’d been working on it for five years, and this book tells all the background of that. And a lot of that was based on Republican philosophy, and that’s always the case with legislation. But here and there you can definitely have an impact.

I could regale you with stories about things that are not in the bill, in part because of research, and I could regale about things that are in the bill despite research. But on the whole, the main thrust of the bill was to increase work and to increase the payback from work, and both of those were based on research. And I think so in that sense research did have an impact. But it had an impact, in part, in large part, by affecting those other factors, like the chairman and the members, the party philosophy and so forth.

BARON: Other questions? Yes?

AUDIENCE (ph): I’m (inaudible) from the Government Accountability Office, and I have a different question. For that Welfare to Work Program, how critical was it when we had an expanding economy, what would have happened if that would have been passed during a recession, and have you controlled for that? And then would have given a bad name to all this other research?

HASKINS: Well, we have, in fact, had a recession, 2001. It wasn’t a deep recession. Come back in three years, we’ll tell you what happens when we had a deep recession. But there’s no -- I carefully said the three biggest factors, I think, welfare reform I think was the most important but you could never prove that with scientific method, but other program changes, especially the earned income tax credit, which Clinton, to his everlasting credit under prompting by David Ellwood expanded quite substantially and then a good economy.

But in 2001 you didn’t have a good economy and you saw the charts. People were still working in an economy where it goes down, but a lot of people still work, a lot of people lose their job, they find another job, so, net, you still had much higher work level than you did in the past among low-income, single women, and their income was much higher than it had been in the past, and child poverty was still way below where it had been before welfare reform.

So there’s no question an economy helps and a bad economy has a negative impact, but work still is a big part of the answer.

BARON: And let me just add that a lot of the random assignment studies that were done in welfare were done over a long period of time from about 1980 through the mid-1990s, which encompassed a number of business cycles, two recessions. And so those results provide, I think, fairly robust evidence across different types of economic conditions and also different areas of the country that had stronger and weaker economies.

RATA (ph): I'm Jimmy Rata (ph) from the University of (inaudible). I have two questions for Ron Haskins. First, on welfare reform, you talked about how the waiver process created an incentive for research and then how the reform created a block grant system. Did that cause the incentive to disappear? Do we still have a vibrant research -- vibrant and funded research approach to even learning from success after reform?

And, secondly, can you address evidence-based efforts around preschool education and Head Start?

HASKINS: As to the first question, yes, I think there's still a lot of motivation, and there's -- I tried to show with that other chart, there's still a lot of work going on, not just in welfare but in lots of other areas, but it's primarily because the federal government pays for it. I think that's the real key, that the federal government -- and this is -- even the Republicans I think could say that this -- paying for research is a legitimate role of government because all the states can profit, the entire nation can profit, and it has in the past, and it will in the future. So I think, yes, there is still a role for research.

If we start cutting the research budgets, which I am afraid will happen -- and in the welfare bill, I can never forget, I had a debate with a Democrat senator once and he was talking about how we needed more research and welfare reform was not based on research. There are about a quarter of a billion dollars in that bill over a seven-year period for various research. There's research on child protection, which produced the biggest study of child protection in the country, $25 million a year, which was actually expanded for the secretary to conduct. I showed you the language using random assignment design.

So I think there clearly has been -- there still is motivation. As long as the federal government keeps paying and we have these great companies like APT (ph) and MDRC and so forth, I think we will definitely get good results, and we can improve our programs if we pay attention to it.

On the preschool programs, I think it's a great example. It's a negative example, the one that I started with. I think the random assignment study, Head Start, conducted by Westat, definitely shows that there's a long ways to go. Head Start is producing very modest impact. It's nothing, nothing at all like Abasadarine (ph) and Perry Preschool.

And so if we really want to have a good preschool program, which I think we have good evidence that we could have good preschool programs, it could make a big difference. Kids would do better in school, they'd be more likely to work, they'd be less likely on welfare, they'd commit fewer crimes, girls would be less likely to get pregnant, I mean a whole range of really impressive outcomes shown by random assignment studies. But when you try to go national, and this is, of course, this is the Achilles heel of all education research, it doesn't work as well, so we need to keep working on it.

BARON: Take one last question.

AUDIENCE (ph): Thank you. My name is (inaudible).

I have a question for Rachel. When you mentioned at the beginning, talking about microfinance, how difficult it is to measure the stats like ambition, I'm actually concerned about when it comes to these kinds of factors, like, for example, the presence of a mentor in social inclusion programs, I mean, I understand that randomized trials help you, in a way, to take care of the selection bias but I guess when it comes to understand if it's a problem, if you can replicate the program, I mean, how do you take care of these kinds of factors?

I'd like to hear a little bit more about this.

GLENNERSTER: OK. So I'm actually involved in doing a random assignment evaluation of microfinance, and the way you do it, which is very hard, which is why there hasn't been one before, is you take different areas and you randomize which areas a microfinance organization will go into. One of the benefits of that -- and then you just survey a random sample of people in the community that had access to microfinance and the community that didn't have access to microfinance or, you know, multiply by 100, because you need something like 100 without and 100 with.

The benefit of that is that you get to see -- not only do you not have the selection problem, you're not comparing motivated women with unmotivated women. You're also looking at the spillover effects on the whole community so you can see what's the impact on the community of having microfinance.

Now, you also asked about mentorship. I mean, there are people who are specifically looking at, for example, you know, if you provide people motivation or you provide them a mentor, how does that impact. So there's a nice study looking at, you know, do you just give information about schooling or do you give information with, kind of, an inspirational example of someone that you can relate to? And, actually, the information about schooling had just as much an impact as the information with mentor. So you can look at those sorts of issues.

In terms of replicability, which has come up a couple of times, I think there were a number of things -- I mean, Jon said earlier that you want to test a program that works, because what's the point of finding something that doesn't work when the program wasn't actually implemented? So there's an important tradeoff here, because sometimes you also don't want to test a program that's so gold-plated and so efficient that it's not replicable. And if you spend as much time as we have been in the development community, you kind of know that a lot of programs are implemented in a half-hearted, not very efficient way. And in a sense, that's what you want to evaluate.

So we sometimes deliberately say, we think there's so much ambiguity about whether even if this program was done incredibly efficiently, we don't know whether it works. We test, sort of, the best possible program. And then if you find it doesn't work, you're like, "OK, then nothing will work." But then other situations we want to test the gold standard version of something in a normal area with not a lot of resources, because that's what's going to be scaled up. And I think that's what's nice about the de-worming project is it's very cheap, it doesn't need special human capital, so you kind of know that you can replicate it.

So on the whole, on that spectrum of tradeoffs, we tend to focus on cheap, easy-to-implement programs, because it's only really worth the effort in doing that evaluation if it's something that we think can be scaled up in a massive scale.

And just to add, people were saying, well, maybe there are two examples of things that work. I gave two examples of things that we found work, but not to end on such a depressing note, there are actually now increasing number of things that we found to be extremely effective, extremely cost effective, some education programs in India, for example, that Pratham is doing, just incredibly effective at increasing test schools, very, very cheap, very little human resources going in so you can replicate them. And they are now -- the Bill and Linda Gates Foundation, the Hewlett Foundation deliberately testing scale up of Pratham's projects so they can be scaled in a very big way.

BARON: OK. Let's please join me in thanking our panel for that excellent set of examples and thoughts.

(APPLAUSE)

Let me end here by again beseeching you to fill out the feedback form.

Second, I want to recognize a couple people. Katy Hoskins has been the -- if you could stand up for a moment -- has been the partner that we've worked with at MCC in pulling this whole event together.

(APPLAUSE)

And also someone who's in this room who really was a key force in MCC's focus on rigorous impact evaluations, helping to make it part of the whole enterprise. It's Delia Welsh who recently left MCC. Could you just stand up for a moment?

(APPLAUSE)

Thank you all for coming

END