



The Future of the World Bank: A CGD Symposium September 23, 2005

Who Watches the Watchdogs? The Role of the World Bank

Pierre Jacquet, Agence Française de Développement

Ruth Levine, Center for Global Development

Lawrence MacDonald, Center for Global Development (moderator)

[TRANSCRIPT PREPARED FROM A TAPE RECORDING]

MR. MacDONALD: Just a quick word about how we're going to proceed. There isn't a coffee break. I believe there is still some coffee in the back. If you want to stand up and stretch, feel free. Ruth and Pierre Jacquet, if you will join me up here, please, we'll go straight into our next panel. So people should feel free to move around but please do so quietly. I'll give you a moment to get settled here.

Our second panel has the rather provocative title, Who Watches the Watchdogs? The Role of The World Bank. And this is really code for a more complex set of issues that have to do with evaluation.

I think many people needed coffee really badly. They probably need to go to the rest room as well. Why don't I give another minute for people to settle down here.

[Pause.]

MR. MacDONALD: Our second panel is Who Watches The Watchdogs? The Role of the World Bank. And that is code for an elaborate and I think complex set of issues that have to do with evaluation.

We have two excellent speakers. I'm going to introduce our second speaker first.

Ruth Levine is a colleague of mine. She's the director of programs and a senior fellow at the Center for Global Development. She is the head of our Global Health Policy Research Network and one of the first outputs of that network was a book that was being published just as I arrived at CDG called A Million Saved, which as many of you know, proven success is a global health--documents 17 successful case studies. Even Bill Easterly agreed that these were successful uses of aid.

If you haven't had a chance to look at it, it's a very moving account and it reflects the passion that Ruth brings to her work in development. She has also been the lead force

behind another Center initiative, making markets for vaccines, which has put forward very specific proposals on advanced market commitments that could create incentives for the private sector to invest in R&D for vaccines for neglected diseases such as malaria and tuberculosis.

It's really my privilege to work with her and I always think if I could be more like Ruth Levine, the world would be a better place. So I'm really pleased that Ruth can be here.

Pierre Jacquet, our first speaker, is the executive director in charge of strategy and the chief economist at the French Development Agency, and he's part of a club of 25 economists, I find this so interesting, who produce a daily editorial for French radio channel and I've often wondered if we couldn't get a club like that going here, and maybe we can have Angelina Jolie be a member of that club.

Pierre, the floor is yours. We look forward to your comments.

MR. JACQUET: Thank you very much. Good morning, everybody. It's a pleasure to be here and I want to thank Nancy and yourself and all your colleagues for this invitation.

Before starting about evaluations, let me just offer one comment about the assertions, and drawing on what Bill said about the double question that we face, first, is what can aid do?, and second, what causes development? It strikes me, unless I'm missing one of his books, that when you look at the current textbooks on economic development, when you look at the different chapters, and even in the index, you will find very little reference to ODA and development aid, and that strikes me as quite a question that we should look at. Why is it that we believe that aid can do something and that in textbooks about development economics, we don't find a mention about development aid?

I leave that with you but I think it illustrates exactly what Bill Easterly told us about the distinct aspects of this question.

In my remarks, I would like to comment on evaluations from my point of view at the French Development Agency, what do we do. It seems to me that our agencies, both multilateral and bilateral, have spent a lot of resources in evaluating the projects and programs.

They have tried to do it generally well, and if there is a problem, we have to analyze it, why. And it seems to me that we have been trained to kill three birds with one stone, and while doing that we have missed the fourth bird, that matters even more.

The three birds are the following. First, it seems that evaluations provide, they play a cognitive role of building knowledge about the countries in which we intervene, about the societies, institutions, their needs, the quality of their governments. In short, that it an empirical field of field knowledge, that helps us build field experience, and

helps our staff become the seasoned officials that can deliver quality development aid in the field.

That is a very important aspect of evaluations. It is important for the institution. It needs to be conducted in house because it's owned by the institution. It is part of a necessary process of self-education.

I feel that critics are generally misplaced about that, because they tend to compare this process of evaluations with what a scientific process of impact evaluation should be, and this clearly is something different and this is of course of fourth bird that we don't kill with our evaluation system, and I'll come back to that in a minute.

The second thing that our evaluations do is to play a management of quality role. Through our current evaluations, our evaluators check, observe results against ex-ante expected ones and are able to monitor the execution of our operations.

What was the objective of the program or project? Was the money used for what it was supposed to? What was the time schedule for disbursement? How to explain any difference between actual and expected. This is an exercise in identification of and check of good practices, and also has to be lodged in house.

The third thing that we try to do with our evaluations is to have them play an accountability role through which our agencies provide an assessment of what they do and communicate it to the outside. And we have a problem then, because I think there is, in the country where this is done, some sort of a mixup between accountability and communication. As it stands, our system of evaluation hardly allows a reasonable rating of what we do. This aspect of evaluation needs to be conducted in a free, independent way, but there may be different methods, including peer review.

Let me comment on the fourth bird that we don't kill, that is, evaluating the impact of our operations. Impact evaluation is a demanding task because it requires both careful observation of direct and indirect results, and careful assessment of attribution of these results to what we do.

Recent progress in scientific impact evaluation methods, and in the literature about these methods, make it now possible to learn a lot about the actual development impact of our operations, and this is a welcome move to developing this kind of approach.

It is welcome for many reasons. First, it allows to better understand what development is about and what works in terms of actions and policies, and what doesn't work and why.

Second, and as a consequence, it will allow more a effective selectivity policy in development agencies as they are able to focus on what works, and to avoid spending money on what doesn't.

There are some caveats, though. First, such impact evaluations are expensive and especially, they have a public good component. That means that they are going to be typically undersupplied, unless there is leadership, coordination, and cooperation to our joint impact evaluations.

That is a system in which many donors take part, each of them contributing and all sharing the knowledge that is produced.

Second, available scientific methods such as random evaluation, but there are many others, will work on some operations and much less on others.

Forecasts, as well as practicality reasons, not everything can undergo scientific impact evaluation. Not everything is amenable to, say, random assignment, like experimentation with drug use. We should therefore start by selecting a few appropriate projects and adopt a forthcoming pragmatic approach. We, at AFD, have started taking part in such a program of impact evaluation in a very specific area, which is not health, not education, but which is a random assignment of parcels of land in which we are going to use a no till technique instead of tilling the land.

And we expect that to be able to demonstrate what the positive impact of no till could be in countries devastated by erosion, for example.

A third caveat I wanted to mention is that even though we need to go ahead with scientific impact evaluation, we also should keep a critical eye on the results from such scientific evaluations.

It is unlikely that any scientific method of evaluation will allow grasping all relevant factors in the interpretation of how a project or program fails in a given context, especially the institutional, human, and societal dimensions. What we do need is a scientifically informed knowledge based on our actions, not a book of recipes about development.

In particular, it is puzzling that we are seemingly putting more time and energy in properly organizing the evaluation of development projects than we seem to be in trying to assess the effectiveness of public policies in our own countries.

There is something bizarre in using developing countries as a laboratory for assessing public policies for development aid, while there is plenty to do in our own countries as well.

There is a second task, a fifth bird, if you wish, that our current evaluation systems doesn't kill, and that is provide a systematic benchmarking or rating of donor agencies' activities, multilateral as well as bilateral.

Such a rating might be useful for two different sets of reasons. First, accountability, and I already mentioned that our current system of evaluation doesn't fulfill that role well.

We need to know whether taxpayer money has been well-used and whether it has served useful purposes from the taxpayers' perspective, and rating of agencies can help that.

Second, let me take a Michael Klein market of aid perspective for one minute. It would be useful to know about the relative performance of donors in order to be able to concentrate resources in the most-performing donors or to think of cooperative ventures between the most performing, not only among public donors but also including private foundations, NGOs, municipalities, and so on.

In a world in which there is increasing competition among donors, the quality of information about these donors is key to efficiency, maybe even more than harmonization.

There are already very interesting initiatives in that respect. SIGET [ph], which is a club of donors involved in micro finance, took a very interesting initiative a couple of years ago, in organizing a peer review of its member activities. Results were discussed in a symposium last year, and identified donors that performed well and others that didn't. That was some sort of a cultural revolution in donors.

Interesting, at least in one case, a donor judged to perform poorly decided to retire from micro finance operations as a result of this peer review. Such an approach obviously could be replicated in other areas.

In a similar vein, the French government has invested in a fund parallel to the global environmental fund, the French global environmental fund. And this fund, before replenishment of its resources, undergoes an audit, both about its operations and its strategies, and again, I think this kind of method should be applied much more systematically.

What can be concluded from these succinct observations about evaluations?

First, agencies still need their own evaluation teams as important knowledge providers from their own perspective and as contributor to quality management. Therefore, we would be mistaken if we presented the necessary move towards scientific independent evaluations as a deep critique of what is being done. This is not the same purpose.

Second, these evaluation teams in our agencies provide little insight into our actual impacts, and however useful the contribution to knowledge is limited to a better understanding of operational constraints and local contexts.

Hence, the need to develop impact evaluations that, because they clearly exhibit public good characteristics, require cooperation among donors and joint action.

As for judgment of performance, this is clearly not a mission for the donor agencies themselves. Their responsibility is to be accountable, that is, to provide all verifiable, accurate, unbiased information on their operations. That is called transparency.

Assessment of performance needs to be totally externalized and shouldn't be even undertaken under a contract directly commissioned by the donors themselves. So there is an interesting question here about how to organize accountability.

Finally, all these dimensions of evaluations are complementary. For effectiveness and efficiency reasons, they should be carefully identified and organized separately.

Some need to be conducted in house, some outside in a comparative peer review or independent manner. We need to consider scientific impact evaluations but we also need to maintain the nonscientific empirical approach that is also part of knowledge building and quality control. Knowledge is not only a scientific good.

Our major shortcomings of evaluations so far is also the insufficient attention given in our agencies, and in the world, I believe, to evaluation results and to the process of organizing the operational feedback from these results to new operations.

Let me conclude very briefly with a few remarks on what The World Bank could do.

From this discussion, I see a number of roles that The World Bank could usefully undertake around both the idea of leadership taken by a big multilateral institution and of the example that it could give to its donor partners.

First, to take on the subject of evaluations and help the community of donors share a doctrine on that important activity.

As a multilateral institution with a clear commitment toward improving aid effectiveness, and researching on development process and results, the Bank is uniquely placed to push a subject higher on donors' agendas.

Second, the Bank, under the initiative of its chief economist, has recently engaged into a broad exercise in scientific impact evaluations, and has proposed to extend such approach to partner donors. My agency is putting together a program along these lines, as I have mentioned already.

It is important that the Bank confirms its leadership role and also acts as an exchange platform about progress achieved by participants throughout the evolution of this program.

Third, the Bank is also uniquely placed to organize a public available knowledge base about the results from impact evaluations. That would serve the public good dimension of impact evaluations. It is also uniquely placed to publicize the results of its own and others' evaluations.

More tentatively, I would support a joint action by donors alongside the SIGET example, to organize a rating of their operation in specific areas. As was exemplified by the SIGET initiative on micro finance, this can work and provide very useful insights about what works, what doesn't, when administered by who.

Here, again, a forthcoming commitment by the Bank, which didn't participate in the Siget peer review exercise, would be useful and welcome. Thank you very much.

[Applause.]

MS. LEVINE: There isn't a lot left to say after that. I think that so many of the points were really very eloquently made, and right on target, and I agree with virtually everything that was presented so far.

Let me just tell you what perspective I'm coming at this question from. I worked at The World Bank for a number of years and designed operations in the health and education sector in Los Angeles and in Eastern Africa.

I participated in the preparation of final project reviews for more than twenty of the Bank's operations in the social sectors.

I spent about two years doing technical reviews of social sector projects at the IDB, which has very similar issues, and I experienced, firsthand, the challenge of trying to incorporate learning into not only the Bank's operations but into the way the Bank functions, from my lowly level, in most instances, as a task manager, and throughout I felt just a deep sense of frustration.

That here was an institution that was able to mobilize and make available billions of dollars for very worthy-sounding projects affecting millions of lives around the world, and yet it was an institution that systematically fails to take advantage of those investments, to learn what actually works to make a difference, so that similar investments in the future could yield better results.

I'll start with an obvious point. To succeed as an institution, The World Bank has to succeed in its main business, and it's main business is financing projects and programs that lead to better economic and social conditions than would have occurred without the projects or programs.

And this is an even higher standard for which the returns are at least a little bit higher than the costs of the programs.

The World Bank doesn't succeed on the basis of whether it remains solvent or gets along well with nongovernmental organizations or keeps employees happy and well fed, or fights corruption at home or abroad.

All of these things may be important, they may be necessary, but they're not at all sufficient for the institution to succeed.

The Bank's success rests on whether it can actually make a difference in the lives of those who are sometimes referred to as the ultimate beneficiaries, better off in some meaningful way, and the obligation incumbent on the Bank to do this is even higher than on bilateral donor agencies because of the nature of the instrument, that these are loans or credits and in some form must be repaid.

So the question is does the Bank do its business in a way that's likely to lead to success? And I'm focusing on the social sectors, health and education cause that's what I know best and that's where I think some of the central evaluation challenges are.

And it's worth keeping in mind that the social sector programs have constituted the lion's share of The World Bank's poverty reduction agenda, and so are really central to what the Bank things it's about.

And this is a focus on what really constitutes success, is as important for programs like poverty reduction credits, which are broad block grants, sort of, or block loan sort of arrangements, as they are for highly projectized arrangements, because what really matters in the end is how the money is spent.

The typical World Bank social sector project is designed with the following inputs, sometimes generated by World Bank staff and consultants and sometimes generated by governments receiving the loan or credit.

First, a very detailed description and analysis of the current bad conditions. Second, guesses about the reasons behind those bad conditions, typically informed by a development economics perspective. Third, some kind of theory of change or argument that says if you make these particular sorts of investments and institutional changes, the world will be a better place in these specific ways. Fewer children will die, more children will go to school and become educated, so they can make a living in the future, and so forth.

And not only will more of these good things happen because of the programs investments in institutional changes, but those good things wouldn't have happened or would not have happened so quickly in the absence of this program. That's the line of argument that is embodied within virtually every program or project.

What's missing, in general, as an input into the design of virtually all of these programs is an actual evidence base to systematically support or refute that theory of change.

Will those particular investments in institutional changes really make a positive difference or do they just sound good?

Have those investments resulted in the desired change before, in the same country, region or elsewhere?

We don't have any systematic information or evidence about that, and that's why Lant Pritchard [ph] in a famous article, said that review meetings at The World Bank are like armies clashing in the night, just sort of arguing without the benefit of light and evidence.

We don't have such information, as Pierre Jacquet mentioned so eloquently, because good impact evaluations are not part of the practice of international development, in general, and there are lots of very good reasons for this.

The technical difficulty, particularly within the social sectors, the public good nature of generating that knowledge. That there are few incentives within the institution to really learn, and I think that anybody who has worked for any period of time at The World Bank or other similar institutions knows that it is virtually impossible to communicate bad news, that is, the project didn't work, up to higher levels of management.

That at each level, that information gets attenuated, gets minimized, and so in the end, virtually all projects end up being given a grade of satisfactory.

And I can tell you from very personal experience, that if you ever rate your project as unsatisfactory, that is the signal that you will tend receive a tremendous workload to produce action plans for the improvement of your project and to document how better things are becoming over time.

So there are tremendous, very practical disincentives to really looking at whether something is working.

And finally, there's a lot of political resistance to impact evaluation because the more evidence that accrues, the less political discretion there can be within the borrowing countries about where the resources can go.

But the fact remains that for most types of programs, a body of scientifically generated evidence about effectiveness is lacking, and Bank task managers are left on their own. The general strategy is to observe that, well, other projects use the same theory of change, and there are a plethora of anecdotes to support it, there's some best practice style documentation and there are some wise and great people who are in accordance with these ideas.

There's a sort of default to eminence-based decision making as opposed to evidence-based decision making.

Fortunately, a lot of people have recognized this problem and I think the observation that this same argument probably occurred 30 years ago is apt here.

Within the Bank, the operations evaluation department makes heroic efforts to squeeze knowledge out of the experiences of projects that are conducted without baseline data, without comparison groups, sometimes without any impact indicators at all.

The chief economist and other colleagues at The World Bank currently are making, again, heroic efforts to try to identify at least a few projects where good impact evaluation methods can be applied.

I think that progress will not be made in this area, solely within the context of The World Bank.

I think that it requires a much broader, collective solution, and a qualitative way in the way we think about evaluating development programs.

I think there are four elements to this solution, and probably more.

One is that impact evaluation efforts should be applied to important questions, enduring questions in development and social development, and we should not expect that every program or project would have an elaborate impact evaluation, and I think the example of the micro finance peer review was a very interesting one.

The second element is the use of evaluation methods that actually yield answers. As Pierre Jacquet mentioned,, there are some true advances in the use of random assignment and other methodologies and I think that those can have an increasing role.

Third, and I think this also echoes the points earlier made. There's a lot of potential to have a collective response which would be consistent with responding to the public good nature of the enterprise, and being able to learn across donors and international financial institutions programs.

And finally, and I think this is really the heart of it, is that I believe that the solution rests on the independence of the impact evaluation, in getting that function largely out of the purview of individual development agencies, not because they can't do it, not because they won't do it perhaps--that's still to be seen--but because of the essential nature of the credibility of the knowledge that is generated, and for better or worse, I think that it's quite hard to imagine that knowledge that is disseminated by The World Bank will be seen as completely credible information. And that may be a controversial point that people wish to take up.

That doesn't, as Pierre Jacquet mentioned, that doesn't in any way suggest that the evaluation activities of myriad sort that are currently going on shouldn't continue and be strengthened. It's simply that the addition, to kill that fourth bird I think requires independent.

And just in closing, I want to say that I think that that's particularly important if there is an increasing move, as I've heard, there's a lot of discussion about toward a more results orientation within the bank, and it may seem contradictory to say, well, there's a greater orientation toward the institution and individuals within getting results, if pay is going to be tied to whether or not you do a good project or not.

Then maybe that fosters better evaluation. But I would argue quite the opposite. I think that if you tie the performance of--or if you tie the career paths of individuals and the welfare of the institution to the specifics of whether or not given projects succeed or don't, then what you have is a distortion of what those evaluation results will yield, and which programs will be evaluated and what the reports will look like.

It always happens, and I think it would just happen in an exaggerated form. So for that reason, I think that yes, push the Bank toward more focus on results but at the same time separate out and strengthen the function of impact evaluation into an independent facility.

So with that, I'll close, and thank you for your attention.

[Applause.]

MR. MacDONALD: [Off-mike] [inaudible] at the beginning of her presentation said she reads nearly everything [inaudible] and I guess the key elements in it, as I understand it, are the independence of the evaluation unit and random assignment, and I heard you mention both of those things.

But do you see both of those as being useful ways to increase knowledge about what works--

MR. JACQUET: Yes. Yes, indeed. I think that in a number of cases, random assignments provide a very interesting method and indeed should be used.

However, I would also stress the fact--but it's [inaudible] agreement, I think we agree on it as well--that there are other methods, and in some cases these other methods will be visible while your random assignment is not, and therefore we should not simply focus on random assignments as if it was the deus ex machina, allowing to solve the problem of our knowledge about impact and results.

So we have new methods, and as Ruth said very well I think, the idea is not to generalize the use of this method about all our operations, because that would be fastidious and in some cases it would be counterproductive, but to pick up a number of subjects in which it matters a lot to know whether programs work or not.

I mean, in education, in health, it's tremendously important to have a good idea of the assessment of some of our programs, and in that case, I think indeed, random assignments and sometimes as a method, provide a very interesting way forward.

Now on the independent, I think--I agree and at the same time, I think that what we need to do is to organize. Independence can come from the multiplicity of actors at a point, and having a group of agencies organize together, full peer review, the process of evaluation, is as valuable from my point of view as trying to set up an independent body, which in fact will never be totally independent because we'll have to pay for it to provide the evaluation and therefore you create a relationship through which this body needs to count on future contracts, and therefore that might distort the evaluation that it provides, and so on.

There are a number of problems that we won't probably be able to solve perfectly. But their method and I think the key is doing it together and the reason why is not necessarily independence but it's [inaudible] dimension, and I would insist on separating the knowledge building aspect of evaluation from the judgmental aspect of it, which one would expect to be able to judge what agencies do. This is something different. It refers to the accountability side of the story which is different from the knowledge base that we need to build. We need both but I would argue that it's best to keep those two tasks a little separate, at least analytically.

MR. MacDONALD: Ruth, Pierre's [off-mike] actually what I'm asking is you made a broad case [inaudible]?

MS. LEVINE: Very briefly, we, at the Center for Global Development, have had a working group of people who care about this topic and are knowledgeable about evolution of social programs, going on for about a year now. Looking at how evaluations have been conducted, looking at the ongoing initiatives, thinking about the nature and causes of the problem and trying to arrive at a proposed solution, or approach to make a qualitative improvement in the situation, the working group has, in general, converged around the nature of the problem of impact evaluation and some of the core elements of a solution in terms of identifying these enduring questions, using appropriate methods, not always random assignment but sometimes, and taking advantage of learning across different agencies.

There are differences of opinion within the working group about this question of independence and the level and type of funding for such evaluations. We currently have what I can call a staff draft of that report, which includes both the analysis and then the quite specific proposal for how to move forward.

It is not intended as a consensus statement of the working group but we are trying to get comment on it, and basically, the core idea--I'll be brief--the core idea is that there would be an impact evaluation facility or club, which donors, multilateral development banks, and developing country governments could belong to on a voluntary basis, and that the club or facility would have the functions of establishing what those sort of core enduring questions to look at are, assisting with the design of impact evaluations, supporting that design, providing the dissemination function that you were talking about, including making data available for reanalysis. I think that's a core feature.

And then, again, there is a question about whether funding would also be allocated through this mechanism for the conduct of impact evaluations and for those of you not on the CGD mailing list, please leave your cards and we'll make sure to get you information about how to access that report.

MR. MacDONALD: I think if my microphone is working now, and I apologize if you couldn't hear me before, we'll now open up to questions. I see one hand. I'm going to be a little stricter than I was last time. These should be brief comments or questions. You, and then there are a couple ladies in line at the microphone. Why don't we take those three, to start.

MS. DENTON: Hazel Denton, John Hopkins University and also consultant to the Bank's evaluation department. I want to pick up on the point that Ruth Levine was making, that there are systematic failures to take advantage of learning. I review health projects and my approach is I pick up the appraisal report, I go through it and I say here, here, and here is where I think they'll have problems. This and this I think will work.

I bat about a 90 percent score when I evaluate myself. The evaluations are thorough, critical and I think independent. Where we're completely lacking is the feedback mechanism to task managers. Why? There is no incentive to learn. There's no extra built in, that congratulates a task manager for saying this is what hasn't worked in the past so this is why I'm doing something differently. Can we start with something very simple, straightforward, and focus on that link?

MR. MacDONALD: Thank you.

MR. PORKA [ph]: Axel Porka, also World Bank private sector development group. I do fully agree with the emphasis on [inaudible] evaluation and in fact we started now with some randomized trials, for the first time, for PSD projects. There still is a very long way to go. I'm somewhat worried that both discussions this morning, and yours, seem to go back to the project level discussion exclusively.

So, to some extent, you get a sense that Bill is arguing that we should only do what we can measure and he'll argue since we can measure projects, let's do projects very soon. My own management will tell me, since we can measure lending volume, do lending volume.

And I would, I guess, also would like to emphasize that there is a kind of public good and knowledge provision function and that maybe the challenge to us should be to better measure what we should do, other countries should do.

That doesn't mean that [inaudible] doing business on the investment climate surveys, and basically assuming, because you measure institutional performance, that we are also able to influence it. But perhaps even providing benchmarking and kind of allowing countries to give themselves orientations as to where to go can be a very important leverage for change and has been, I think, working reasonably well with the

entry indicators which we used in IDA 14. So again, just a plea, that in all this discussion about projects, that the knowledge side doesn't get completely lost. Thanks.

MS. RUBIN: Hi. I'm Nilamini [?] Rubin. I work for the Senate Foreign Relations Committee. Senator Lugar recently introduced development bank reform legislation which included text on evaluations, and in the drafting process the most significant feedback we got on the evaluations piece was the cost.

When Ruth Levine was talking about the challenges to impact evaluation, she mentioned technical incentives and political. I was wondering if you could speak to the challenge of cost. Who should bear the cost? Should it be the Bank or the project, the recipient? What is the general percentage of the project cost that'd be the evaluation cost?

And really, how significantly does the cost of measuring outweigh the cost of not measuring?

MR. MacDONALD: Excellent question. Thanks very much. I think we'll take those next two, if you can make them quick, and then we'll turn back to our panel for responses.

MS. : Hi. I'm a graduate student in education and my question is what kind of independent evaluation clubs or that might include donors or recipients--what kind of outside--ar there agencies in place right now that could serve as these independent evaluation agencies, or what kind of research is being done on who would perform those evaluations, and are there effects being made to create those right now? Like who would do this evaluation?

MR. MacDONALD: Who will watch the watchdogs? Right. And the last comment or question, please.

MS. ALEXANDER: Nancy Alexander, Citizens Network on Essential Services. I have two questions. Both of them are about tracking whole loan instruments, and Ruth, you did an effective job of describing of how evaluations get watered down as they go up the line. But when the evaluations relate to a whole entire group of loans, like decentralization loans, OECD did a great evaluation of decentralization, showing the many ways in which it's not effective in reducing poverty.

There was an annual report on development effectiveness last year, that talked about the whole category of community-driven development funds and social funds, and how they were most often captured by elites. Antonio Astash [ph] did a really wonderful report on public/private partnerships and infrastructure over the last eight years, talking about many of the problems in terms of fiscal drain reaching the poor, cream skimming, and each of these instruments, even though there have been these critical evaluations, the papers for the development committee don't reflect any of the learning, and in fact they're being scaled up in the Africa action plan.

So what do you do about--

[Tape change, Tape 2A.]

MS. ALEXANDER: [In progress.]--the new World Bank evaluation on capacity building in Africa says that the Bank has spent a billion dollars in capacity building over the last few years in Africa and none of it has been routinely tracked, monitored, or even evaluated. So that's an enormous deficiency on the part of the institution, but there don't seem to be any mechanisms to rectify those problems.

MR. MacDONALD: Thanks very much.

I actually liked it last time, ending with audience questions, since this is a symposium and part of an ongoing dialogue. So I'm going to ask Ruth and Pierre to respond as you wish and that will also be your last comment. And then I'll take, if there are additional quick comments or questions, we'll have them in the transcript and hope that, as these papers come out, they will help you think through the papers that you will be contributing to the symposium volume.

Who would like to go first?

MR. JACQUET: Ruth wants the last word.

MR. MacDONALD: We'll let you come back after Ruth. We'll give you an extra last word, since she had the last word in the presentation.

MR. JACQUET: I was joking.

I think these are very important questions, all of them, really. First, there is really a difficulty with what was mentioned by the first speaker there, that there is no incentive to learn from evaluation results. I think this really points to process. That is why we need to think about process in our agencies. And part of what we do in the evaluation teams right now, it's also to help us check on that process and be sure that the results of past evaluations are available in the conception stage of the forthcoming projects or programs. That is something very important, that shows, actually, that the future is not only in scientific impact evaluations; it is also in the process and management in our agencies.

Second, indeed I totally agree with the fact that we shouldn't only do what we can measure and we shouldn't only do what we know has worked in the past. There is a responsibility that we have in a development agency, which is to innovate to produce more knowledge. And in a way, in my view, development aid should be much more than it is--something like venture capital. We commit funds and we take risks. And the outcome may be failure, but from failure we should learn. Failure might be desirable, in a way, because we learn from our failures.

We don't quite do that in our agencies. It seems to me that, over the past years--and this is a worrying trend--we are increasingly tempted to lessen the risk that we take and, in a way, to check so much about whether we are going to succeed, there is a good chance to succeed, whether what we do meets a number of new standards, and so on. We are bureaucratizing the decision process. This runs against efficiency. So there is a right balance to find between these different aspects. So again, when we speak about evaluations, no dogma. Let's just be very pragmatic about what we need to progress on.

Cost of evaluations, that's a very good point. I think that we should spend more time discussing whether the cost of scientific impact evaluations is actually more than paid for by the results that we expect from those evaluations. And indeed, there is a question mark there. Although from the very low level of where we start, I think we make no mistake whatsoever in committing resources at least to launch some scientific impact evaluations. But from the point of view of an individual agency like mine in France, the cost of true scientific impact evaluations is gigantic. So we need, really, to think about how to get resources to help us doing that.

I will leave the comment on independent evaluation.

Oh yes, independent evaluations on education or what else, I think--no, from my knowledge there aren't any existing outside agencies and there shouldn't be, probably, because what matters about scientific impact evaluation is a partnership between existing agencies and the academic community. And probably governments and local governments and local stakeholders as well. That is what a comprehensive scientific evaluation should be about.

Last question, this was the most difficult aspect of it because there are programs that are very difficult to evaluate. And what you mentioned about capacity building and a number of other broad programs belong to that category. That reminds me of what I feel as a teacher and as a professor in my school. The way we try to evaluate our teaching is simply have students fill questionnaires of satisfaction. This is not impact evaluation. And I'm very happy that my students are satisfied, but I'm a bit frustrated because, when I'm conscious--I should drink more, probably--but when I'm conscious I realize that this is not exactly impact, but the usefulness of what I teach. Somehow, I don't think that much has been done to solve the difficulty.

Well, we need to accept that the world is not perfect and that we as development agencies in a [inaudible] world and that we won't be able, probably, to know everything that we do.

Thank you.

MR. MacDONALD: Ruth?

MS. LEVINE: Just very briefly I'll take up a couple of the points. On cost, I think there is no way, at the moment at least. I don't know how to quantify precisely whether or not

the costs of doing an impact evaluation are greater or less than the value of the knowledge that is generated. I agree that we're extremely, extremely far from any sort of frontier where I think lots of hard choices in that regard need to be made. We're really operating at a very, very low level currently.

I think we can take some instruction from the experiences of domestic social programs in the U.S., where, for example, in the area of job training and welfare, the very careful work that has been done to evaluate those programs has yielded tremendous insights that I think, really, observers across the political spectrum and from multiple different disciplines would agree have really been to the benefit of the effectiveness of those programs. So I think we can take some comfort in that, that the money would be well-spent. And we do have in our proposal an estimate of what a, kind of, reasonable amount would be to generate results for 10 major evaluations per year, so I will invite you to look at that. It's not a mind-blowing number.

And then on the question, I believe it was the first one, about the lack of feedback between evaluation results, particularly from the Operations [inaudible] Department to [inaudible] managers, I think that is a separate but tremendously important problem. I don't know what the solution to that is. Maybe it's something about this greater orientation to results. It's my observation that in many cases task managers actually depend almost exclusively on their very own experience in the field and with operations that they have been involved with prior, rather than on any systematic understanding about what the results have been for other task managers in other countries or even in the same country.

And as a result, and because task managers tend to design a program, get rewarded for that for the signature, and then move on--and so they often don't see how the programs they designed played out. For that reason, the Bank actually ends up being less than the sum of its parts, which is perhaps the most depressing observation I could make. But I think there is a solution to that.

MR. MacDONALD: We've got three more fascinating panels. I'm going to double back on my words. I'm not going to ask for more questions from the floor. Instead, I'm going to give Pierre Jacquet the final word here. Then we'll go straight into our next panel.

However, if you do have something you wanted to say, I know that both Pierre and Ruth would like very much to hear it, or you can write it down and e-mail it either to them or to the Center for Global Development. And while you're there, sign up for our e-mail newsletter so you can track the progress of the evaluation gap initiative.

Pierre, the last word is yours.

MR. JACQUET: Thank you. No, I didn't want to have the last word, but listening to what Ruth was saying, I have another comment to make, which is that we should reward staff officials not about their results but about the capacity to learn from failures. We should certainly not rule out failures. Failures are part of a normal learning process. So

everything that was said about tying reward to results I think to me is very dangerous, not only because of great distortions but also because it creates the wrong incentives. The incentive is not to get good results always; it is to learn from failures, and to qualify failures. So it's a different perspective that I think matters a lot.

Thank you.

MR. MacDONALD: It will be interesting to see in Paul Wolfowitz's speech tomorrow in his push for results if he's going to tie results directly to reward. You've described the risks, I think, in a very provocative and accurate way.

We'll stretch for a moment. Coffee, if you want it. If our next two panelists--that would be Masood Ahmed and John Taylor--please join me here on the stage.

Could I have a round of applause, please, for these excellent panelists.

[Applause.]