

Very preliminary
Please do not quote without the authors' permission

Making aid work

Abhijit V. Banerjee and Ruimin He
M.I.T
October 15, 2003

One would think that giving away money ought to be easy. After all, there are so many who need it so badly and we have a pretty good idea of where they are. Yet, rather remarkably, we seem to have arrived at a point where more or less everyone agrees that aid-giving is not working the way it ought to.

A part of the problem is that most people are not actually giving away money---or rather, while they often give away money, the ultimate recipients do not get it as a straight gift of money. Somewhere along the chain an NGO or a government is responsible for turning this money into schools, or hospitals or roads, or whatever the people are supposed to get.

What makes giving away money hard? The bias against just giving people money stems, in part, from the feeling that the best use of the money may not be to spend it on consumption. This is plausible, indeed likely: There is some relatively convincing evidence that many people do not invest as much in their businesses and their children's human capital as the rewards to such investment seem to warrant. What is less clear is why people could not be relied on to make the right investments on their own---in which case it would be enough to hand them the money. One possible reason may have to do with the lack of self-control: It may be too tempting, especially for poor people, to spend the money on something they need right away. The incompleteness of the intra-family "contract" is another reason not to trust the family with the money: Parents may put too little weight on improving their children's earning capacity, because they do not expect to share in their children's prosperity. And of course, people may not know what is good for them.

There was a time when many of these kinds of arguments could not be made among respectable economists---it was seen as a transgression against the freedom of the individual and bad economics to boot. Now the pendulum has swung to the other end---it is now an item of faith in the development community that no one should be giving away money. It is not clear, what, if any, evidence lies behind this shared conviction. Certainly no one has done the experiment of showering large gifts of money on poor people in poor countries and then following them to see what they do with the money and what happens afterwards.

A very different kind of argument for giving people goods rather than money, comes from the fear that if you are seen handing out money, even those who have enough of their own may want to pretend to be needy. The advantage of providing public services rather than money is that the non-poor may not want them enough to make it worth their while to simulate poverty. The very rich in the US, after all, choose to pay for their children to go to private schools, even though their children are entitled to go to public schools at no charge, simply because they feel that the public education system is not quite up to the standards they aspire to. This saves the government the cost of finding teachers for these extra children. On the other hand, these parents do not hesitate to claim any tax deductions that they may be entitled to which suggests that if the government was giving away money instead of schooling, the rich would be in the queue with everyone else.

It is however not all obvious that one cannot give away money without opening the floodgates. After all, the rich value their time: Making it necessary to queue up in order to collect the money should discourage those who really do not need it.

The broader point here is not to deny that giving away money has significant disadvantages, but to emphasize that we know very little about how serious these disadvantages might be. In particular, are these costs necessarily large enough to outweigh the significant costs of trying to give away anything other than money?

Delivering goods and services to the poor

If you do want to give away money but still think it is worth trying to help the poor, you would have to give them things---roads, schools, banks, hospitals, fertilizer. Giving away things is more work than giving away money for the simple reason that someone needs to produce them: roads have to be built, teachers hired and trained, fertilizer produced and hospitals kept in good repair. It would be simple if it were just a matter of paying for the roads, fertilizer etc, but that is just the first step. Then we would have to make sure that the roads are built to the required standards, that the teachers teach, the fertilizer gets to the right people, that the process of delivery is not subverted by corruption and/or bureaucratic incompetence....

What is perhaps even more difficult is to be reasonably confident that the money is being spent on things that are really worth getting, in the usual sense of being at least as good as any other way of spending the same amount of money. After all, there are many ways one could spend money on promoting education---one could build more buildings, hire more teachers, provide free textbooks, distribute free uniforms, put flip-charts in classrooms, set up computer labs in every school, provide a bonus for teachers who teach well, serve hot meals in school and much more. Indeed every one of the many interventions listed here (and many more) has been tried somewhere in the world just in the last few years. This diversity reflects, in part, differing needs, but often it is just ignorance: Lacking a clear sense of what works, well-meaning donors will choose what their intuition suggests, even though it may be very different from what the donor next door believes. Both seem to believe that they are right.

There is also no guarantee that it makes sense to spend the money on directly promoting education. The fact that people are not getting an education on their own might reflect the lack of jobs for those who have an education. Pushing out more graduates who will not find jobs either, may actually be counter-productive, because it might reinforce the lack of faith in the value of education. The best way to promote education may be to create jobs.

Or we may even want to look beyond education. Perhaps one should invest in health and leave education to private initiative. Or forget about both and go for fertilizers. How should we decide where to go?

It is therefore no surprise that the process of helping the poor by giving them access to goods and services is fraught with difficulty. There needs to be a system for picking the right project and a system for making sure that the project is carried out as it should be. And for figuring out how much people are getting out of it, and whether it continues to be what they need or want.

For this donor will need to get involved in the process of decision-making and delivery at the ground level, though the exact nature of the involvement can vary substantially: the actual production of the good is usually contracted out, though even this is not always the case. Both the process of delivery and its impact has to be assessed, though once again there is a choice between doing it yourself and contracting it out. And the broader strategy needs to be worked out, based, one presumes, on knowledge of the situation on the ground. This might involve consulting local experts. Even carrying out some new research as a prelude to the intervention is not out of the question.

In all of this the donor will typically work with one or more local organizations, be it government departments or NGOs. Some of these local partners may have their own sources of funding, which will allow the donor to leverage its resources.

Whatever the exact strategy, whether the donor does the monitoring or preliminary research or whether these are contracted out, these involve very substantial expenditures over and above the direct cost of delivering the good (or service) to the ultimate beneficiaries. To get a sense of how large these expenses might be, we note that between the years 1996 to 2001, the World Bank administrative budget per year averaged US\$1401 million¹, while the total World Bank IBRD and IDA loan per year² averaged US\$15615 million and US\$6154 million respectively. The Operations Evaluation Department of the Bank calculates ex-post Economic Rate of Returns for the projects that it evaluates³, but only for certain sectors⁴, and finds that the median revised⁵ economic rate of return for both the IBRD and IDA/blend lending operations for Fiscal Year 1996-2001 exits⁶ was about 20%. The aggregate economic return on the Bank's portfolio was 20% of \$15.615bn + \$6.154bn, i.e. about \$4.400bn. It follows that the benefits from World Bank lending are lower by almost a third because of administrative costs.

Of course, as we will discuss later, we should not expect to entirely avoid these costs by switching to donating money rather than goods. For now, however, the relevant point is that these costs are large.

¹ Appendix 5, IBRD/IDA Appendices, World Bank Annual Report, 2000; and Appendix 1, IBRD/IDA Appendices, World Bank Annual Report, 2002.

² 1999 and 2001 World Bank Annual Report, Page 10 and Page 26 respectively.

³ From Table 12, Statistical Appendix, 2001 Annual Review of Development Effectiveness.

⁴ Agriculture, Electric Power & Other Energy, Environment, Mining, Oil & Gas, PSD/Industry, Telecommunications & Informatics, Transportation, Urban Development, and Water Supply & Sanitation.

⁵ Ex-post.

⁶ "Date that the project leaves the World Bank's active portfolio": Notes to Table 12, Statistical Appendix, 2001 Annual Review of Development Effectiveness.

Evaluating donor effectiveness

Given everything that they spend on the design and management of aid projects, are donors getting what they hoped for? The short answer is, unfortunately: we do not know. A part of the problem is data: As Appendix 1 shows, most of the larger “public” donor organizations (as against private foundations), do evaluate their projects, but they usually stop short of a summary quantitative assessment of the social impact of the project, such as a rate of return. Of the donors listed in that Appendix, only the World Bank reports rates of returns and then only for certain sectors---education, health and nutrition for example, are left out. This reflects, in part, doubts about whether it makes sense to try to reduce the many dimensions of a project outcome to a single rate of return. In part it also reflects the inherent difficulties of coming up with a rate of return---how do you come up with the right counterfactual which tells you what would have happened in the absence of the project.

What some of these donors do instead is to assess overall project performance combining both process evaluation (“the right number of schools were built”) and impact evaluation (“the children’s test scores improved”), on a set scale⁷ (Appendix 1).⁸ However this evaluation is often carried out by those who are also involved in the implementation of the program, making it somewhat hard to know what to make of the results. Finally, many of these organizations do not allow the public access to their assessments. Of the eight organizations listed in Appendix 1, only two, the World Bank and the Asian Development Bank, have their projects assessed by a formally independent organization, report project level assessments on a set scale and allow the public access to the assessment results. These are the only two therefore that offer the possibility of delving deeper.

This is very far from ideal. The World Bank is not just any donor organization: It is probably the most visible organization of this class, with all the constraints that come from being in the public eye. It is formally responsible to those who provide its financing, which are the governments of a handful of rich countries. Moreover, it only gives loans (albeit on very attractive terms). It is also an organization that attracts and employs many of the best minds thinking about development today. Perhaps most importantly, it sees itself as a leader in the efforts to promote development in the world: this probably mean that its projects need to be evaluated not just in terms of what they directly achieve but also in terms of how they shape efforts outside the Bank to promote development. In a previous paper, Banerjee and He (2003), we try to evaluate the Bank’s achievement as a leader and conclude that there is no evidence that it is having a huge impact outside bank. But we also argue that the Bank is ideally placed to take this leadership role and that it is important that it does so, which obviously implies that we must take it seriously in evaluating the Bank’s performance.

⁷ Such as satisfactory/unsatisfactory or successful/unsuccessful.

⁸ Others, like USAID, just report their assessment of how well the country is doing.

The ADB is also quite special, being the one multilateral funding organization that has close ties with the Japanese government. We may expect it to have been influenced by the Japanese government's rather distinct view of economic policy.

Given that the World Bank and ADB evaluations are on different scales and the evaluators have potentially different standards, there is no point in trying to compare these two organizations. One could, of course, take their assessments at face value: In 2002, the Operations Evaluation Department of the World Bank wrote that "at the project level, the outcomes of Bank-financed projects continue to improve", and more than 60% of all projects evaluated each year since 1990 have had satisfactory outcomes⁹; while the Asian Development Bank wrote that the proportion of successful projects/programs (by year of completion) has been more than 50% since 1997, and has been trending upwards since 1989¹⁰.

The problem is that we do not really know what to make of the scale they use---what does it, at some absolute level, mean to say that the World Bank's Operations Evaluation Department feels that the project was satisfactory. How much of this assessment reflects, for example, what they expected (which we do not know) rather than some objective that we all share. It seems safer therefore to compare assessments *across different projects or sets of projects for the same organization*.

World Bank or ADB projects vary considerably in the degree of their involvement, as measured by the share of the project financed by these organizations. If we assume that putting more money into the project reflects a greater commitment to the cause, we can use this ratio as a measure of the bank's priorities.

What can we say about the World Bank's priorities? For each project approved (i.e. launched) between 1994 and 2001, the World Bank reports the share of World Bank funding in total funding for that project.¹¹ For the period 1987-2001, we also have the evaluation of projects by sector¹², averaged over 3 year periods (1987-90, 1990-93, 1994-97 and 1998-2001)¹³. We label these four periods 1, 2, 3 and 4. We then regress the share

⁹ Page xiii, Executive Summary, 2002 Annual Review of Development Effectiveness.

¹⁰ Page 37, Annual Review of Evaluation Activities in 2002.

¹¹ Percentage figures from the Annual Reports of the World Bank: Summaries of Projects Approved for IBRD, IDA, and Trust Funds in each Fiscal year. For each project, the following information are provided: Country name, sector, brief project description, World Bank contribution, Total project cost.

¹² Sectors were determined by the classifications in Annex 1 of the 1999 Annual Review of Development Effectiveness (which conveniently classifies the historical performance data into the 1990-1993, 1994-1997 periods). From these sectors, the "Social" and "Environmental" sectors were dropped because zero/one evaluation was done between 1990 and 1993. We do not have 1987-1990 data for "Mining", "Multi-sector" and "Public Sector Management" due to a change in sector classification.

¹³ The 1987-1990 performance data is derived by collating information from Table 1.2 of "Evaluation Results for 1991" and Appendix Table 7 of "Project Performance Results for 1987". The years 1987 and 1990 here refer to the year that the project was evaluated; hence they correspond to projects that had slightly earlier exit fiscal year groups. The 1990-1993 and 1994-1997 performance data is derived from Annex 1, Table 1 of the 1999 Annual Review of Development Effectiveness, and correspond to projects

of World Bank funding in a particular project approved in period t on the average evaluation in period $t-2$ of the sector that it belongs to [$Prevperf$], and the improvement of its evaluation between $t-1$ and $t-2$ [$Diffperf$]. We control for fixed differences across sectors, countries and periods and cluster errors by sector.

The results in Column 1 of Table 1 show that when a sector's performance improves, projects in that sector get a higher fraction of their financing from the Bank (the $DiffPerf$ coefficient is positive). But it also helps to start at a low base (conditional on the same degree of improvement, projects in sectors that started with a worse record get more money: the coefficient on $Prevperf$ is negative), which immediately implies that if two sectors have shown the same improvement, the one that is doing worse will get more money from the Bank.

Table 1: OLS Regressions on Project Selection

	World Bank	ADB [including multi/others]	ADB [excluding multi/others]
	(1)	(2)	(3)
<i>PrevPerf</i>	-.002553* [.0011]	-.006767* [.0033]	.0004253 [.0047]
<i>DiffPerf</i>	.001852* [.00051]	-.004361* [.0014]	-.001184 [.0020]
<i>N</i>	1513	519	468
<i>Adjusted R²</i>	.29	.33	.39

Dependent variable: % of individual project that is funded by the World Bank.

Significant at the 5% level of significance. Robust standard errors reported.

A very similar pattern emerges when we look at the total amount of money allocated to each sector. For four sectors, Agriculture, Finance, Technical Assistance, and Water and Sanitation¹⁴ the World Bank provides data from 1974¹⁵ to 1993¹⁶ for every block of three years¹⁷ on the sector outcome measured by OED evaluation¹⁸, and the log of total expenditure on that sector.¹⁹ Using this data set, we regress²⁰ [$\log(expenditure_t)$] on [$\log(expenditure_{t-1})$], [$outcome_{t-1}$], and [$outcome_t - outcome_{t-1}$], correcting for period and sector effects. The results are shown in column 1 of table 2. They show that increased spending for a particular sector is associated with an improving trend in sector performance over

that had exit fiscal years within the respective ranges. The 1998-2001 data is derived from collating information from the Annex 1s of both the 1999 and 2001 Annual Reviews of Development Effectiveness.

¹⁴ These are the 4 selected sectors in Annex Table 1.32, 1995 Evaluation Results Volume 2: Annex.

¹⁵ Data starts from 1974.

¹⁶ The Technical Assistance category was dropped in the 1994 Annual Report.

¹⁷ With the exception of the 1992-1993 year.

¹⁸ Annex Table 1.32, 1995 Evaluation Results Volume 2: Annex.

¹⁹ Annual Reports of the World Bank. It's way too much effort to collect data on the percentage funding for all these years.

²⁰ The Arellano-Bond linear, dynamic panel data estimator is used.

the immediate past. Once again, having started from a lower initial level of performance helps, but the coefficient is not significant in this case.²¹

Table 2: Cross-sectional Panel Regressions on Project Selection

	World Bank (1)	ADB (2)
$\text{Log}(\text{expenditure}_{t-1})$.7677* [.028]	.1523 [.18]
outcome_{t-1}	-.002256 [.0026]	.000925 [.0089]
$\text{outcome}_t - \text{outcome}_{t-1}$.001594* [.00075]	-.0004837 [.0073]
Number of obs	16	45
Number of groups	4	7

Dependent variable: $\text{Log}(\text{expenditure}_t)$

* Significant at the 5% level of significance. Robust standard errors reported.

When we do the same exercise for the ADB, we get very different results. Both past performance and improvement in performance seems to have a negative impact on the allocation of its funding (column 2 of table 1). However this result is very sensitive to the inclusion of the “Multi-sector/Others” sector, where there were only 2 evaluations between 86 and 89 (both successful). Once this sector is dropped, both the past level and the improvement in the level become insignificant (column 3 of table 1). All that matters then are fixed cross-country differences and perhaps cross-sector differences.²²

The results for the ADB seem very consistent with the view that its priorities are largely set by high level decisions and are not particularly subject to any short term influences. This may be a good thing because it insulates the organization against the influence of fads and internal political shifts, but it clearly also prevents the organization from learning from its experience.

The case of World Bank is more complicated. It clearly does not give priority to sectors that have been performing the best over the immediate past, which is what, under the (possibly brave) assumption that past performance is a reasonable index of what we might expect in the immediate future, would have been the way to maximize immediate impact. But it does favor the sectors that have been improving the fastest. One way to rationalize this may be to assume that the World Bank sees itself as a leader in the development community: As a leader, it would make sense for it to try to promote those sectors where the potential for improvement is the highest, rather than those where, on current record, the possibility of success is the highest. Sectors that have been improving

²¹ Here, unlike in the results in the previous table, we are using the gap in performance between t and $t-1$ rather than that between $t-1$ and $t-2$, because of data limitations in the case of the World Bank. For the ADB we could use the gap between $t-1$ and $t-2$ ---the results are very similar.

²² Both for the ADB and the World Bank, the sector dummies as a group are significant at the 1% level (based on their joint F-Statistic) as are the country dummies, though the sector dummies become insignificant for the ADB when we drop the multi-sector/others category.

fast over the past few years but still have some distance to go, may therefore be exactly the sectors the Bank would want to favor.

It is however possible to take a more cynical view of the same evidence. In this view, the Bank is excessively influenced by shifts in current fashions in development thinking. The reason, in this view, that we see the Bank reacting to improvements, is that these improvements are correlated with shifts in fashions---after all, one reason something becomes a fashion is that there was a breakthrough in an area that had hitherto been doing badly.

The question, in the end, comes down to whether the projects that are being given priority are doing what they were intended to do. One way of looking at this is to look at the correlation between the fraction of planned project financing that was to come from the Bank and the performance of the project according to the Bank's evaluators, after controlling for fixed differences across sectors and countries, the length of the project, the year when it was approved, and the year of the evaluation. The results are shown in Column 1 of Table 3. As already reported in Banerjee and He (2003), there is a negative and significant correlation between the priority that Bank originally gave the project (measured by the fraction of financing that was supposed to come from the Bank) and its performance. Bank favored projects seem to do worse from an *ex ante* point of view, relative to other projects in the same sector.

This negative relation goes away if we replaced the share of planned cost that the Bank was supposed to pay for with the share of actual cost (Column 3 of Table 3). Basically if projects are going really badly, the Bank cancels its promised contribution to them. But even with the help of this corrective procedure the correlation between performance and funding remains insignificant: Being a Bank priority does not help you perform better, even after the cancellations.

The negative relation between project funding and project outcomes is also apparent when we compare across sectors. We regress [% of World Bank funding] on [sector outcome] and [length of project], correcting for country, approval year, and closing year fixed effects, and clustering by sector. As reported in Column 2 of Table 3, the coefficient on Sector Outcome is significant at the 12% level. Once again, this negative relation goes away when we replace the share of planned cost that was supposed to be paid for by the Bank with the share of actual cost (Column 4 of Table 3).

We repeat this exercise for the ADB, using projects evaluated²³ between 1997 and 2002. Here percentage funding is defined as the [Loan Amount *approved*]/[Expected Project cost] in Columns 5 and 6 of Table 3, and [Loan Amount disbursed]/[Actual Project cost] in Columns 7 and 8 of Table 3. We find that the outcome rating is not significantly affected by being given a priority (under either definition), irrespective of whether we compare across sectors or within. The one exception is in the actual cost share regression, where the outcome variable is negative and significant at the 10% level.

²³ Data from the respective Annual Review of Evaluation Activities "Profile of Evaluations in xxxx".

Table 3: On effectiveness of fund allocation

	World Bank (planned)		World Bank (actual)		ADB (planned)		ADB (actual)	
	Within (1)	Betwn (2)	Within (3)	Betwn (4)	Within (5)	Betwn (6)	Within (7)	Betwn (8)
<i>Outcome</i>	-.177** [.062]	-.554 [.36]	.0173 [.024]	-.1520 [.22]	-.0345 [.032]	-0.183 [.17]	-.056* [.031]	-.161 [.17]
<i>Length</i>	-.0197 [.016]	-.0327 [.017]	-.030** [.010]	-.0368** [.012]	.0166 [.011]	.008598 [.0076]	.00835 [.012]	.0015 [.010]
<i>n</i>	664	664	664	664	137	137	136	136
<i>Adj. R²</i>	.19	.15	.25	.18	.53	.44	.50	.38

Dependent Variable: Percentage of funding of individual projects. The within regressions control for sector dummies, country dummies, year of approval dummies and year of closing dummies. The between regressions include country dummies, year of approval dummies and year of closing dummies.

** Significant at the 5% level of significance. Robust standard errors reported.

* Significant at the 10% level of significance.. Robust standard error reported.

The lack of a positive correlation between funding and performance, in the case of the ADB, seem unsurprising, given everything else we seen about the way they target (or rather, do not target). In the case of the World Bank, these results are consistent with the view that the Bank is faddish. They can also explained by assuming that the bank is particularly bad at running its projects.

On the other hand the results are also quite consistent with any view that has the World Bank playing the role of a leader and prioritizing projects that others, more focused on immediate impact, would not choose. After all, we already knew that they have not given priority to the sectors that had performed best in the past---all that this evidence really shows is that this is true within sectors as well. Of course, as already pointed out, it is not clear that the World Bank is particularly effective as a leader, and to the extent that we take this as given, we may still want it to focus more on its rather limited direct impact.

In the end there is very little that is reliable that we can say about donor performance. The most we can say is that we found no prima facie evidence of great effectiveness.

What limits effectiveness?

Donor organizations are in many ways very much like other organizations, and they share many of the standard organizational constraints. Organizations like the World Bank obviously need many people to act and take decisions on their behalf, and there is nothing to guarantee that they have the right incentives. In particular in an organization that lives by doing projects and making loans, no one achieves prominence by rejecting projects and refusing loans. For this reason alone, most people on the implementation side of the Bank, either in their sector departments or in their country missions, ought to be somewhat biased in favor of making something happen. Add to this the fact that they are the ones who deal with the potential recipients, and we have a recipe for a degree of over enthusiasm and irresponsible lending.

There is also, obviously, the possibility of ideological conflict within the organization. We recently saw a public example of such a conflict in the World Bank, which ultimately lead to the resignation of the person in charge of the 2000-01 World Development Report. There must also be other fights that have less to do with ideology than with personalities and individual ambitions. All those involved in these fights must be tempted to use the power to sanction projects to help their supporters and punish their enemies.

There is also the fact that being a donor organization involved in development makes you someone that many people want to influence---ranging from the US government to NGOs hostile to the US government. These pressures are probably easier to resist for an organization that has an explicit ideology or an acknowledged political master----like the Salvation Army or USAID---than a nominally apolitical organization like UNDP or the Bank. For such organizations the challenge is often in maintaining their reputation for being open to many different views while continuing to make the right choices about the allocation of funds. Resisting the temptation to placate the different sides by conceding some of their less merited demands, must be challenge for these organizations.

However the biggest problem, in our view, is the lack of an *explicit* scientific basis for their decision making. A recent World Bank publication---Empowerment and Poverty Reduction: A Sourcebook²⁴ provides an excellent case-study of the way decisions get made in the Bank.

The Sourcebook is meant to be a catalogue of what, according to the Bank, are the right strategies for poverty reduction. These are also, we presume, strategies that the Bank is prepared to put its money into. It provides a very long list of recommended projects, which include: computer kiosks in villages; cell phones for rent in rural areas; scholarships targeted towards girls who go to secondary school; schooling voucher programs for poor children; joint forest management programs; water users groups; citizen report cards for public services; participatory poverty assessments; internet access for tiny firms; land titling; legal reform; micro-credit based on group lending; and many others.

While many of these are surely good ideas, the book does not tell us which of these have been subject to a randomized evaluation or even a non-randomized evaluation based on a convincing quasi-experimental design. In fact, to the best of our knowledge, only one of these strategies has been so evaluated in a LDC context---schooling vouchers for poor students---and that was because the Colombian voucher program had built-in random allocation.²⁵ Yet it receives no more weight than any of the other programs.

Indeed most of these programs are recommended on the basis of very little hard evidence. Legal reform, for example, is justified by asserting that “The extent to which a society is law-bound affects its national income as well as its level of literacy and infant mortality”.

²⁴ Narayanan, Deepa, ed. *Empowerment and poverty reduction: A sourcebook*. Washington DC: World Bank, 2000.

²⁵ See Angrist et al (2001).

This may be true, but the available evidence, which comes from cross-country correlations, hardly warrants such a confident recommendation.

And some programs, it seems, no amount of negative evidence can stop: Our favorite example is the *Gyandoot* program in Madhya Pradesh in India which provides computer kiosks in rural areas. The Sourcebook acknowledges that this project was hit hard by lack of electricity and poor connectivity and that “Currently only a few of the Kiosks have proved to be commercially viable”. It then goes on to say, entirely without irony: “Following the success of the initiative....(pp 80)”

Making aid work

The limited role of evidence in the decision to promote or continue particular initiatives, obviously favors those who have reasons, political, ideological or purely private, for not wanting the best projects to be chosen. This must exacerbate the agency problems within these organizations. Nevertheless we do not feel that this is the main reason why these organizations have not moved to a different standard.²⁶ From our experience dealing with senior officials in donor organizations, we feel that they genuinely believe that there is no practical alternative to the current system. The key to making aid work lies in convincing these people that there is a workable alternative.

Their objections to the idea that policy should be evidence-based typically fall into one of two categories. First, there is the fear that requiring that every initiative be justified in terms of hard evidence, will bias decisions in favor of what is measurable and easy to evaluate. Second, there is the view that there is simply not enough hard evidence around, so that basing policy on such evidence would be tantamount to inaction.²⁷

We feel that both these concerns are substantially exaggerated. In terms of the first concern, we obviously recognize that there are things like macro-policy that are very hard to evaluate properly. The problem is once something is big enough (“currency boards”, “democracy”) there is going to be no way to know what would have happened in its absence. And yet, there are clear examples of policies that, most people would agree, make very little sense (“over-valued fixed exchange rates”, “a pension plan that is headed for bankruptcy”, for example). Helping governments in their efforts to get out of these indefensible policy positions, has to be a good use of donor money.

Once we go beyond these “obvious” interventions, there is still a range where we can expect to make a evidence-based decision. These are typically macro interventions that allow for the possibility of a limited micro evaluation. For example, while decentralizing power is a macro reform, we could learn a lot about it by looking at the impact of an initial pilot, where the reform is only implemented for certain areas.²⁸ When we go beyond these, it is still useful to try to make use of the best available evidence: Evidence

²⁶ This is the position taken by Pritchett ()

²⁷ This is, in effect, what Stanley Fischer, a former Chief Economist of the Bank, put it, while commenting on our previous paper at the meetings of the American Economic Association

²⁸ These would be chosen randomly from some larger set.

can at least help in choosing a model whose assumptions correspond better to the world. Obviously how far the donor is prepared to go down this evidence “quality-ladder” will depend on the donor: What is key is that she has a sense of what she is giving up---the fact that there may be other projects out there for which we have much more reliable (and reassuring) evidence.

The other side of this same concern is that requiring evidence discriminates against projects that promote less measurable outcomes---such as female empowerment. Now it is true that historically the focus of economic measurement has been on concepts like consumption and income rather than empowerment, but when Chattopadhyay and Duflo (2003)²⁹, needed a measure of female empowerment in the context of public action in Indian villages, they used the fraction of questions asked by women in village meetings. While this is not perfect, it is not obviously worse than using income to measure well-being, as we regularly do. We are therefore optimistic that once we commit ourselves to measurement, the interaction of the donors and the evaluators will generate a range of good measures of most things that are relevant.

To address the other main concern---that basing action on evidence will lead to paralysis---we carried out the following crude but useful exercise. We began by searching for interventions that have been subjected to randomized evaluations and appear to work. To come up with a list, we asked researchers in the Bureau for Research in Economic Analysis of Development (BREAD) for references and used summary papers by Kremer³⁰, Knowles & Behrman³¹, De Cock et al³², and the Working Group 5 of the Commission of Macroeconomics and Health³³ as starting points for a literature search. In addition, a web-based search was used. From these we deliberately left out regulations, such as Tobacco Taxes, Bans on Tobacco, etc. The Table of Interventions in Appendix 2 lists all the papers that were eventually included in our list. It is meant to cover every category of micro intervention that we would find which has been subjected to an evaluation.

From these papers, we highlight a subset of them as programs that, based on currently available evidence, appear to be programs that should be scaled up globally. Three criteria were used to make this choice: the program must be randomly placed, sustainable without a strong intervention by the researchers³⁴ and must show a significant positive impact on at least one of the initially chosen objectives. If two experiments show

²⁹ Duflo, Ester and Chattopadhyay, Raghabendra, “Women as Policy Makers: Evidence from a India wide Randomized Experiment”, NBER working paper #8615, December 2001

³⁰ Kremer, Michael, ["Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons,"](#) *American Economic Review Papers and Proceedings*, May 2003, pp 102-106.

³¹ James C. Knowles, Jere R. Behrman “Assessing the Economic Returns to Investing in Youths in Developing Countries”, mimeo

³² De Cock, K.M. et al, “Prevention of mother-to-child HIV transmission in resource-poor countries: translating research into policy and practice.” *JAMA*, 283:1175:1182.

³³ Commission of Macroeconomics and Health, Working Group 5, http://www.cmhealth.org/cmh_papers&reports.htm#Working%20Group%205

³⁴ We feel that field workers ensuring that villagers comply with their daily supplement dosage constitutes overly strong intervention.

different results, we favored the one that has the better experimental design. We left out all programs that simply gave away money, but included programs (like school vouchers) that give people money that can only be used for a specific purpose. Finally, we were not quite sure of how to deal with a program like PROGRESA in Mexico, which makes an income transfer to mothers who send their children to school. The problem is that we have no idea of how much we would need to pay to get children into schools, and how much is a pure gift over and above that. So we treat it as a separate category.

We are not at all suggesting that the interventions listed are the only ones that work. Neither do we suggest that the programs that we list but decided not to scale up, are necessarily worthless. It is entirely possible (indeed we hope that it is true) that there are many other interventions that do work but are not included in our final list either because no randomized evaluation has been done on them or because we are not aware of the randomized evaluation. Other programs that were dropped because of overly strong research intervention might have worked with less intervention---we simply do not know. Yet others may really work in their non-experimental variant, but not in the form they were implemented during the experiment. Finally, there are probably many interventions yet to be thought of that have the potential to change the world.³⁵

Given this list of “successful” interventions, we ask the question: how much will it crudely cost to scale them up to a global level? Our definition of global covers only Low Income Countries (LICs), unless stated otherwise. We take the population of each country to be the average of the current population and the projected 2015 population. 2015 is chosen because it is when the Millennium Development Goals are meant to be achieved and many of our programs could be ways to achieve the goals.

The way we calculate costs is to take a point estimate of the per person cost for each program and then adjust it for each country. This adjustment takes the form of a Purchasing Power Parity conversion for intervention in goods (such as supplements), and a Gross Domestic Product per capita conversion for services (such as teachers). Within each country, the size of the targeted population depends on the country’s demographics. These corrections are explained in Table 4 under each project highlighted (for estimates from other studies, the sources of these studies are provided). The cost spreadsheet is available online³⁶. Figures are normalized to year 2000 US\$.

The results are shown in table 4. Our calculations show that a recurring annual expense of about \$11 billion could be justified by the hard evidence we already have, without including PROGRESA. In addition, if we accept the (non-experimental) results in Cutler et al. showing that the health returns to improved water supply are enormous, we should

³⁵ We also recognize that researchers do not always have control over the experimental design, and in any case, that the research might have been done with different objectives in mind (for example, researchers might be interested in knowing the impact of supplements on growth if it could be *assumed* that compliance was not going to be a problem). Inclusion in the highlighted subset does not reflect any opinion of the quality of the research or the researcher. It merely reflects the suitability of the research for our purposes.

³⁶ <http://web.mit.edu/ruimin/www/whatworks/whatworksest.xls>

consider investing in water supply infrastructure. This will cost an additional (one-time) \$15 billion.

\$11 billion is a substantial amount. It is more than the World Bank gives out as IDA loans (the main form of World Bank aid) in a given year (as previously mentioned, an average of US\$6154 million between 1996 and 2001). If we add to this what countries absolutely need to make essential macro adjustments (of the kind discussed above) there may be very little left from the total available donor money.

To the extent that there is still money, it can be used to provide humanitarian aid: There are people in the world who are dying because they do not have enough money to buy food or medicines. Giving them money (or food, or medicines) may not promote development but it is hard to imagine that it would not be good thing.

Indeed once we decide that we are willing to make cash gifts to people, we could make that gift conditional on the recipients fulfilling certain conditions such as sending their children to school. This is what PROGRESA does.

It is true that this does not deal with most of the objections against giving away money: The targeting problem was solved in Mexico by using the bureaucracy to make sure that money goes to the right person, which is part of what makes the program costly. On the other hand, it is not clear that perfect targeting is worth the effort, given that most people in developing countries are actually quite poor. It may be better to set up rather lax criteria for eligibility so that only the rich are ineligible, and then to randomly check claimants and impose harsh punishments on those who are caught cheating. It is usually quite easy to identify the rich in poor countries (ownership of a car, for example, might be used).

It is also true that we do not eliminate the possibility that the money may just go into consumption. But even this is less of a problem than it might appear: A number of studies (including those based on the PROGRESA experiment) have shown that more money in the hands of the female members of the family in poor countries does translate into better nutrition for children and better healthcare.

Clearly being open to the idea of giving away money will make it much easier to find things to do. Just scaling up PROGRESA, by our calculations, will cost \$22 billion a year, which pushes up total annual expenditure on “good” programs to \$33 billion.

Conclusion: A new challenge for the millennium

We live in an age of aid pessimism. There is a strong, if rarely completely articulated, presumption that aid can at best help people survive—it cannot promote development. The US government’s new initiative, the Millennium Challenge Account (MCA), is based on the idea that the whole idea of aid giving needs to be rethought: In particular, they want to tie aid to country performance—only countries that pursue economic policies that the US government approves of will be eligible for aid from this account.

The premise is that aid has not working because the policy environment is not right. While it is clear that this is a problem---there countries where the risk of the money ending up in a government official's pocket are substantial---the thrust our argument is that the way the money is planned to be spent is also a very big problem, but a problem whose source lies in the way the donor organizations function. Combined with the fact that many of the world neediest live in the ones that will not make on the MCA list and that we expect the incentive effects of the MCA to be minimal, this suggests to us that the MCA approach amounts to abandoning a large part of the worlds poorest for no fault of their own. A more effective and less unfair challenge may be to try to see if it is possible to design projects that work in the countries that have the biggest problems. If we could make it work, we would not only help those who need it the most, but what is perhaps even more valuable, we will raise expectations and build hope.

Table 4: Program costs

PROGRAM (Recurring annual expenditure)		R	Cost (US\$₂₀₀₀ Millions)
Education			
Balkashi (Community remedial program)	Abhijit Banerjee et. al, "Remedying Education: Evidence from Two Randomized Experiments in India", MIT mimeo, July 2003	Y	667
	Proportion of students that need remedial scaled by the (1 – primary school completion rate). Cost data from www.pratham.org/reports , and in general, country data from the 2002 Human Development Index.		
Universal education at a 40 : 1 student : pupil ratio	Angrist, Joshua and Victor Lavy (1999), "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," Quarterly Journal of Economics, Vol 114 (2), pp. 533-575. ³⁷	Y	1540
	Proportion of students that need program scaled by (1 – current primary school enrollment rate).		
School Inputs (Uniforms and textbooks)	P. Glewwe, M. Kremer, S. Moulin, "Decentralization: A cautionary tale", Harvard University mimeo, Mar 2003	Y	2270
	In the situation that all primary-school aged children go to school (for 6 years). Transportation costs ignored. Cost data from Glewwe et al. (2003)		
Vouchers	Joshua Angrist et al., "Vouchers for Private Schooling in Columbia: Evidence from an Randomized Natural Experiment" <i>American Economic Review</i> , Dec 2002, pp. 1535-1558;	Y	1399
	Assuming that everyone is sufficiently motivated to achieve satisfactory performance, hence qualifying for the vouchers; and ignoring general equilibrium effects due to the resultant increased in private school fees. Unit cost used is the increased public educational expenditure per lottery winner, given in Angrist et al.		
Rewarding parents for sending children to school (Involves giving away money)	Jere Behrman, Pilali Segupta and Petra Todd , "Progressing Through Progres: An Impact Assessment of a School Subsidy Experiment", Penn Institute for Economic Research, Working Paper 01-033, April 2001	Y	22390
	Assuming that with sufficient subsidies, everyone wants to send their children to school, hence we give the subsidy to everyone. Unit cost data from Behrman et al.		
Supplementation			
Iron	Gustavo Bobonis, Edward Miguel, Charu Sharma, "Iron Supplementation and Early Childhood Development: A Randomized Evaluation in India", Pratham USA, May 2003	Y	280
	2 to 6 years old. Cost data from Miguel and Bobonis, private communication.		

³⁷ Angrist and Lavy's (1999) paper does not deal with a randomized experiment.

Albendazole (De-worming)	Edward Miguel and Michael Kremer , “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities”, forthcoming <i>Econometrica</i> , Feb 2003	Y	29
	2 to 6 years old. Cost data from Miguel and Bobonis, private communication.		
Transportation and Delivery cost via the Pratham Delhi Health Program model	See for eg. Gustavo Bobonis, Edward Miguel, Charu Sharma, “Iron Supplementation and Early Childhood Development: A Randomized Evaluation in India”, Pratham USA, May 2003	Y	220
	We use Miguel and Bobonis’s cost estimates for Delhi and scale up from there.		
Vitamin A	West KP Jr et al. “Double blind, cluster randomized trial of low dose supplementation with vitamin A or beta carotene on mortality related to pregnancy in Nepal”. The NNIPS-2 Study group, <i>BMJ</i> , 1999, 318:570-575	Y	71
	0 to 20 years old. Cost data from Pandav, C.S., K. Anand et al (1998) "Cost of vitamin A and iron supplementation to "at risk" population", <i>Indian J Pediatr</i> 65(6): 849-56		
Iodine	Claudine Cobra et al, “Infant Survival Is Improved by Oral Iodine Supplementation”, <i>Journal of Nutrition</i> , Vol. 127(4), Apr 97, pp. 574-578	Y	140
	0 to 6 years old. Cost data from IDDNewsletter, Vol 14 Number 3 (1998).		
	No handle on the transportation cost for the Vitamin A and Iodine interventions. Presumably some of it can be covered by the Fixed costs from the Pratham program.		
HIV			
Condom provision	Allen, S. et al, “Confidential HIV testing and condom promotion in Africa. Impact on HIV and gonorrhea rates.” <i>JAMA</i> , 268:3338-3343	Y	135
Improving STD management	Grosskurth, H. et al, “Impact of improved treatment of sexually transmitted diseases on HIV infection in rural Tanzania: randomized controlled trial”. <i>Lancet</i> , 346:530-536. (1995)	Y	428
Voluntary counseling and testing	Coates et al, “Efficacy of voluntary HIV-1 counseling and testing in individuals and couples in Kenya, Tanzania, and Trinidad: a randomized trial”. The Voluntary HIV-1 Counseling and Testing Efficacy Study Group. <i>Lancet</i> , 356:103-112 (2000)	Y	116
Prophylaxis for opportunistic infections	Mwinga A. et al, “Twice weekly tuberculosis preventive therapy in HIV infection in Zambia.” <i>AIDS</i> 1998; 12:2447-57.	Y	40
Short course zidovudine regime	Shaffer N et al, “Short-course zidovudine for perinatal HIV-1 transmission in Bangkok, Thailand: a randomized controlled trial.” <i>Lancet</i> , 1999:353:-773-780.	Y	4
	For the HIV interventions - Main data source: Kumaranayake L, Watts C. (2000) "Costs of Scaling HIV program Activities to a National Level for Sub-Saharan Africa: Methods and Estimates". Prepared for the World		

	Bank. We scale her numbers by the ratio of the SSA cost to the LIC cost as given in the Commission of Macroeconomics and Health's Working Group 5: Paper 19 report.		
Malaria: Spraying	Rowland M, et al. "Indoor residual spraying with alphacypermethrin controls malaria in Pakistan: a community-randomized trial", <i>Tropical Medicine and International Health</i> , 2000, Vol 5(7): 472-481.	Y	1480
	Malaria assumed to be prevalent in all the LICs. Cost data from http://bvs.insp.mx/componen/svirtual/ppriori/04/0401/arti.htm		
Fertilizer	Esther Duflo and Michael Kremer, "Understanding Technological Choices: Fertilizers in Western Kenya", power-point slides, mimeo	Y	1848
	We assume that a country is fertilizer deficient if both the fertilizer consumption and the cereal yield are low. Cost data from http://www.ifdc.org/PDF_Files/AFM_OCT01.pdf		
Vaccination	Hoke CH et al, "Protection against Japanese encephalitis by inactivated vaccines", N Engl J Med. 1988 Sep 8;319(10):608-14. Katherine L O'Brien et al, "Efficacy and safety of seven-valent conjugate pneumococcal vaccine in American Indian children: group randomised trial", Volume 362 Issue 9381 Page 355. Irene Pérez-Schael et al, "Efficacy of the Rhesus Rotavirus-Based Quadrivalent Vaccine in Infants and Young Children in Venezuela", 1997, Volume 337(17): 1181-1189.	Y	507
	Derived from the additional funding needs of the Vaccine Fund Strategic Plan from 2001 to 2011.		
TOTAL (excl. urban water provision and School attendance subsidies)			11177 million ≈ 11 billion
TOTAL (excl. urban water provision)			33567 ≈ 34 billion
Urban water provision	David Culter and Grant Miller, "Clean water measures in American History", power-point slides, private communication.	N	73
	Water availability and cheapest cost intervention data chiefly from the Global Water and Sanitation Assessment, 2000 Report.		
TOTAL (incl. urban water provision)			33639 million ≈ 34 billion

One-off Infrastructure Expenditure		R	One time cost (US\$ ₂₀₀₀ Million)
Urban water construction	David Culter and Grant Miller, “Clean water measures in American History”, power-point slides, private communication.	N	4450
Rural water construction		N	10900
	Water availability and cheapest cost intervention data chiefly from the Global Water and Sanitation Assessment, 2000 Report.		
Total water construction			15350 million ≈ 15 billion

R: Randomized

Appendix 1: Evaluation Organizations

Organization (Evaluation Office website)	Evaluation Office	Year	% of projects independently evaluated	Level of evaluation detail available online for independent evaluations	Rating scale for independent evaluations	Evaluation categories for independent evaluations
National Organizations						
Department for International Development DFID http://62.189.42.51/DFIDstage/policiesandpriorities/files/ev_home.htm	Evaluation Department (EVD)	1968 ⁱ onwards	Very few: An average of 9 projects were independently evaluated annually between 1993-1999, ⁱⁱ Only 25% of bilateral expenditure is covered by a completion report done by program managers ⁱⁱⁱ , “there is no system of independent verification, although one has been mooted”, and “more significantly, PCRs are not used” ^{iv} .	All evaluation reports and their summaries are either online or can be ordered online.	Mostly prose-based, because independent evaluations are carried out by different evaluators, no standardized rating scale/category exists.	
U.S. Agency for International Development USAID http://www.doe.org/partners/eval.cfm	Bureau for Policy and Program co- ordination ^v , Center for Development Information and Evaluation	Before 1995 ^{vi}	All projects were evaluated ^{vii} , but how many are independent?			
		After 1995 to 2000	Regarded poorly. See footnote. ^{viii}			
		After 2000 ^{ix}	Searching through the USAID evaluation website shows that most project evaluations continue to be institutionally co-authored by the USAID mission to a specific country, although CDIE has some evaluations that do not involve the country mission as a co-author. ^x	Individual project evaluations. Mostly joint authored with country mission, raising questions about independence. Annual Report uses country macro-economic values as performance indicators. ^{xi}	Qualitative measures used even in its performance and accountability report. ^{xii}	
Regional Development Banks						
African Development Bank. AfDB http://www.afdb.org/about_adb/OPEV.htm	Operations Evaluation Office	1964(est.) – 2001				
	Operations Evaluation Department (OPEV) ^{xiii}	2001 onwards ^{xiv}	Annual Reviews of Development Effectiveness starting in 2003. ^{xv} Annual Review of Results of Operations Evaluation apparently published, according to http://www.afdb.org/about_adb/OPEV_evaluation_guidelines.htm however it’s not available online.	Project evaluations prior to 1997 are available at the DAC website. ^{xvi} Qualitative abstracts of Project Performance Audit Reports from 1994 to 1997 available online or by request to the web-master. However it is not clear if these were independent evaluations.	- ARDE 2003 not currently available online yet. -	
Asian Development Bank ADB http://www.adb.org/Evaluation/	Operations Evaluation Department	1974 onwards ^{xvii}	40% evaluated ^{xviii}	Individual project reports (from 1995 onwards) ^{xix} . Quantitative panel data grouped by sector and region available. ^{xx} Summary tables collating evaluation details per project evaluated. ^{xxi}	Overall project rating: Highly Successful, Successful, Partly Successful, Unsuccessful ^{xxii}	Relevance, Effectiveness, Efficiency, Sustainability, Institutional Development (REESI) ^{xxiii}
European Bank for	Project Evaluation Department	1991 (year of establishment) onwards ^{xxiv}	44% (1991 to 1998) ^{xxv}	Examples of successful and less successful projects. Time series	Transition impact: High, Medium, Low,	Impact on market. Overall,

Reconstruction and Development EBRD http://www.ebrd.org/projects/eval/index.htm				quantitative data of performance for all projects grouped together. ^{xxxvi} Annual Evaluation Overview Reports will be available online from this year onwards, in 2-3 weeks time from 9 Sept 2003. 40 project summaries and lessons learnt (each 1 page long) representing “a cross-section of EBRD investment operations” published online. Previous year overview reports and all of the other Project Completion Reports are internal documents. ^{xxxvii}	None, Negative. Overall performance: Highly Successful, Successful, Partly Successful, Unsuccessful ^{xxxviii}	Financial, and Environmental performance ^{xxxix}
Inter-American Development Bank IADB http://www.iadb.org/cont/evo/evo_eng.htm	Evolve over time. (i) Group of Three Controllers, (ii) Office of External Review and Evaluation and the Operations Evaluation Office, (iii) Office of Evaluation ^{xxx}	1959 – 1998	Seldom undertaken, even then, only by borrower. ^{xxxi} Exact figures unknown, but we find that no ex-post evaluation was carried out for Mexico throughout the 1990s ^{xxxii} , which gives cause for worry.			
	Office of Evaluation and Oversight ^{xxxiii} (OVE)	1999 onwards ^{xxxiv}	6 to 7 country programs evaluated each year (at least from 1999 to 2001). However, the lack of ex-post program evaluation in the 1990s means that countries are evaluated on macro-performance. The presence of multiple donors makes assigning credit hard. ^{xxxv} Very few project evaluations are done independently. ^{xxxvi}	Country program evaluations, each covering over a ten year period. ^{xxxvii} No cross-country or cross-sector comparisons.	Quantitative figures, but these are based on project evaluations that were primarily self-evaluated, ^{xxxviii} and sometimes based on the most recent Project Performance Monitoring Reports (as opposed to the Project Completion Reports which were seldom done ^{xxxix})	Relevance, coherence, efficiency and effectiveness ^{xl}
International Organizations						
International Monetary Fund IMF	Office of Internal Audit and Inspection	Before 1996	Insignificant numbers of external evaluation			
	Evaluation Group of Executive Directors ^{xli}	Between 1996 to 2000	Trial run which covered a range of topics and were of different scale. ^{xlii}			

http://www.imf.org/external/np/ico/index.htm	Independent Evaluation Office (EVO)	After 2000	At capacity, 5 projects to be undertaken each year, including both country specific cases ^{xliii} and broader thematic questions ^{xliiv} .	Commitment to promptly publish all reports, unless under exceptional circumstances ^{xliv} No cross-sector (for example, comparing the performance of capital account crisis intervention versus the pro-longed use of IMF resources) or time series comparisons available yet.	While not explicit, key cross-country macro-economic indicators are provided, and implicitly used as indicators of performance. Within the 2 reports published thus far, there is an attempt to distinguish between different levels of "success".	As at October 15 2003, 3 reports have been published. ^{xlvi} These have looked at surveillance, program effectiveness and IMF governance. The medium program is given in its annual report. ^{xlvii}
United Nations Development Program UNDP http://www.undp.org/evaluation/	Office of Evaluation and Strategic Planning (OESP) ^{xlviii}	Before 1996	For programs <US\$1 million, country managers decide if project should be evaluated. Mandatory evaluations for projects >US\$1 million ^{xlix} . Even then, the compliance for mandatory evaluations was less than 80%. ^l			
	Evaluation Office (EO)	1996-1999				
	Evaluation Office (EO)	After 1999	Requirement that all projects over US\$1 million be evaluated is abolished. ^{li} Decentralized evaluation process, with country managers selecting the evaluation teams and designs the term of reference. ^{lii} The EO concentrates on independent, country-level Assessment of Development Results (5 to 10 countries a year), and sector-thematic level evaluations. ^{liii}	Individual project reports available online. ^{liv} Quantitative time series data for each performance indicator ^{lv} (all projects lumped together), and cross-sector (for projects evaluated from 1999 to 2000) performance impact indicators. ^{lvi}	Yes/Partial/No and Significant/Satisfactory/Poor ^{lvii}	Typically covers Relevance, Performance and Success (Impact, Sustainability, and Contribution to capacity building) ^{lviii} . However, this criterion is different from the previous year, which raises some questions. ^{lix}
World Bank http://www.worldbank.org/evaluation/	Operations Evaluation Department	1973 ^{lx} onwards	Independent evaluation at the Country, Sector and Project levels. Project Performance Assessment Reports for 25 percent of all completed projects. ^{lxi}	Individual Project reports. ^{lxii} Quantitative panel data grouped by sector and region available. ^{lxiii} Summary tables collating evaluation details per project evaluated. ^{lxiv}	Outcome: Highly Satisfactory (Sat), Sat, Moderately Sat, Moderately UnSat, UnSat, Highly UnSat. Sustainability: Highly Likely, Likely, Highly Unlikely, Unlikely Institutional Development: Substantial, Modest, Negligible. ^{lxv}	Outcome, Sustainability, Institutional Development. ^{lxvi}

We only look at the level of evaluation detail available online for the latest incarnation of the evaluation office.

Appendix 2: Table of Interventions

Intervention	E	P	H	Evaluation cited	S	R	Benefits	Incl. cost?
Water & Sanitation								
Clean water through a water container with a cover and a sprout			X	Lee Roberts et al, "Keeping clean water clean in a Malawi refugee camp: a randomized intervention trial", <i>Bulletin of the World Health Organization</i> , 2001, 79(4), pp. 280-287.	Y	Y	Reduced diarrhea	No.
				Self-sustainable because Malawian field worker used, and "this [how to use the bucket] educational message generally took less than one minute and was never reinforced or restated during the study". Not scaled up because it was carried out in a specific instance of a refugee camp that had experienced a cholera outbreak.				
Latrine provision			X	P Emerson et al, "Effect of fly control and pit latrine provision on trachoma in preschool age Gambian children", Proceedings: Tenth International Symposium on Human Chlamydial Infections	Y	Y	Reduced trachoma (But not significant)	No. Effect not significant
				Self-sustainable because "pit latrines were constructed by the Gambian Department for Community Development, using the 'Gambian Improved Household Pit' design", and they demonstrated the capacity of providing every household with the latrines post-experiment.				
Latrine provision			X	Daniels DL et al, "A case-control study of the impact of improved sanitation on diarrhoea morbidity in Lesotho", <i>Bull World Health Organ.</i> 1990;68(4):455-63	Y	N	Reduced diarrhea	No. Not random.
				Selection bias: Ownership depends on constituency, and distance to the recruiting health facility.				
Latrine provision			X	Esrey SA. "Waste, waste, and well-being: a multicountry study", <i>Am J Epidemiol</i> , 1996:143(6): 608-623	Y	N	Reduced diarrhea, taller and heavier children	No. Not random
				Potential non-random location of latrines.				
Education for water-sanitation behavior	X		X	Stanton BF, "An educational intervention for altering water-sanitation behaviors to reduce childhood diarrhea in urban Bangladesh. II. A randomized trial to assess the impact of the intervention on hygienic behaviors and rates of diarrhea", <i>Am J Epidemiol.</i> 1987 Feb;125(2):292-301.	Y	Y	Reduced diarrhea	Yes. (We assume that if everyone goes to school, the schools will do the education)
Historical clean water interventions in America			X	David Culter and Grant Miller, "Clean water measures in American History", power-point slides, private communication.	Y	N	Clean water reduces mortality	Yes (although not random).
				Attempts to deal with potential endogenous placement.				
Health-care services				Summary: H. Gelband, S. Stansfield, "The Evidence Base for Interventions to Reduce Under Five Mortality in Low and Middle-income Countries", CMH Working Paper 9.				
Midwife services			X	Damien Walker et al., "An economic analysis of midwifery training programmes in South Kalimantan, Indonesia", <i>Bulletin of the World Health Organization</i>	Y	N	Better skills	No. Not random.

			2002(80), pp. 47-55				
			Not random due to participation selection bias.				
Midwife services		X	Elizabeth Frankenberg and Duncan Thomas, "Women's health and pregnancy outcomes: Do services make a difference?", <i>Demography</i> , 38(2), May 2001, pp. 253-265	Y	N	Increased women Body Mass Index	No. Not random
			Actual program, but selective non-random midwife placement.				
Social Marketing of reproductive health services			"Assessing the Economic Returns to Investing in Youths in Developing Countries" – James C. Knowles and Jere R. Behrman				
Reduction in antenatal care		X	Munjanja SP, Lindmark G, Nystrom L. "Randomized controlled of a reduced-visits programme of antenatal care in Harare, Zimbabwe." <i>Lancet</i> , 1996:348:364-369	Y	Y	No effect on maternal/fetal outcomes	Cost reduction, not included
			Actual reduction in number of visits, clearly sustainable.				
Reduction in antenatal care		X	Villar J et al, "WHO antenatal care randomized trial for the evaluation of a new model of routine antenatal care", <i>Lancet</i> 2001;357:1551-1564	Y	Y	No effect on maternal/fetal outcomes	Cost reduction, not included
			Actual reduction in number of visits, clearly sustainable.				
Home-based neonatal care		X	Bang AT et al, "Effect of home-based neonatal care and management of sepsis on neonatal mortality: field trial in rural India", <i>Lancet</i> , 1999, 354:1955-1961	Y	N	Advert 1 death per 18 neonates cared for	No.
			Control villages were not randomly chosen: Villages were used as controls because suitable women could not be found, or population was less than 300.				
Nutritional Supplements			Summary: James C. Knowles, Jere R. Behrman "Assessing the Economic Returns to Investing in Youths in Developing Countries", mimeo and L. Nemer, H. Gelband, P. Jha, "The Evidence Base for Interventions to Reduce Malnutrition in Children Under Five and School-age Children in Low and Middle-Income Countries", CMH Working Paper 11				
De-worming drugs	X	X	Edward Miguel and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities", forthcoming <i>Econometrica</i> , Feb 2003	Y	Y	Improved health and school participation, even in neighborhood schools	Yes.
			Self-sustainable because "medical treatment was delivered to schools by Kenya Ministry of Health public health nurses and ICS public health officers" and this was done within the community setting (we see some children not participating).				
Iron supplementation	X	X	Gustavo Bobonis, Edward Miguel, Charu Sharma, "Iron Supplementation and Early Childhood Development: A Randomized Evaluation in India", Pratham USA, May 2003	Y	Y	Increased participation, reduced absenteeism	Yes.
			Existing pre-school network used.				
Zinc and Iron supplementation		X	Torbjorn Lind et al, "A community-based randomized controlled trial of iron and zinc supplementation in Indonesian infants: interactions between iron and zinc", <i>American Journal of Clinical Nutrition</i> , 2003, 77(4), pp.883-90	N	Y	Yes, but interaction effects need to be understood	No. Not program
			Not self-sustainable because of overtly heavy fieldworker involvement (ensured that subjects complied with daily dosage).				
Oral Iodized Oil		X	Claudine Cobra et al, "Infant Survival Is Improved by	Y	Y	Improved	Yes

supplementation			Oral Iodine Supplementation”, <i>Journal of Nutrition</i> , Vol. 127(4), Apr 97, pp. 574-578			infant survival	
			Added to the ongoing EPI program.				
Vitamin A supplementation		X	Suharno D et al, “Supplementation with vitamin A and iron for nutritional anemia in pregnant women in West Java, Indonesia.” <i>Lancet</i> , 1993,342:1325-38	N	Y	Decreased mortality	No. Not program
			Not self-sustainable because of strict supervision.				
Vitamin A supplementation		X	West KP Jr et al. “Double blind, cluster randomized trial of low dose supplementation with vitamin A or beta carotene on mortality related to pregnancy in Nepal”. The NNIPS-2 Study group, <i>BMJ</i> , 1999, 318:570-575	Y	Y	Decreased pregnancy related mortality	Yes
			Self-sustainable because local provision, and less supervision.				
Supplements for lactating women			Anne Tinker, Kathleen Finn, and Joanne Epp, “Improving Women’s Health: Issues and Interventions”, Washington, DC: The World Bank, 2000	Y	N	Reduced low weight babies	No. Not random
			Supplements were provided daily on a <i>volunteer</i> basis at a centralized location, i.e. there was no compulsion to consume it.				
Folate supplementation (Literature review)		X	Mahomed K. “Folate supplementation in pregnancy (Cochrane Review)”. In: <i>The Cochrane Library</i> , Issue 2, 2000 Oxford: Update Software	-	-	Not enough evidence to evaluate effectiveness on clinical outcomes	No.
Anti-platelet supplementation (Literature review)		X	Knight M et al, “Antiplatelet agents for preventing and treating pre-eclampsia (Cochrane Review). In: <i>The Cochrane Library</i> , Issue 2, 2000. Oxford: Update software, 2000	-	-	Timing and dosage knowledge insufficient	No
Supplemental feeding (Literature review)		X	Rush D. “Nutrition and maternal mortality in the developing world”. <i>Am J Clin Nutr</i> , 2000, 72(suppl):212S-240S	-	-	Insufficient knowledge to decide if nutritional supplements are good overall.	No.
Tuberculosis			Summary: M. W. Borgdorff, K. Floyd, J. F. Broekmans, “Interventions to reduce tuberculosis mortality and transmission in low and middle-income countries: effectiveness, cost-effectiveness, and constraints to scaling up”				
Short-course chemotherapy		X	China Tuberculosis Control Collaboration, “Results of directly observed short-course chemotherapy in 112,842 Chinese patients with smear-positive tuberculosis.” <i>Lancet</i> , 1996; 347:358-362.	Y	N	Increased cure rates	No. Not randomized.
BCG		X	Tuberculosis Research Center (ICMR), Chennai. Fifteen year follow up of trial of BCG vaccines in south India for tuberculosis prevention”, <i>Indian J Med Res.</i> 1999; 110:56-69.	Y	Y	No effect of BCG on TB	No. No effect.
Preventive therapy		X	Mwinga A. et al, “Twice weekly tuberculosis preventive therapy in HIV infection in Zambia.” <i>AIDS</i> 1998; 12:2447-57.	Y	Y	Drugs reduced TB infection for HIV-infected people in	Yes (under HIV interventions)

							Zambia	
Vaccinations			X	Here we look for evidence for the new vaccines that the Global Alliance for Vaccines and Immunization is trying to introduce. (http://www.vaccinefund.org/en/publications/strategic_plan_2002_eng.pdf). There is a strong evidence base for the more established basic vaccines. Here I assume that all the vaccines can be implemented as programs, due to the success of the Expanded Program for Immunization.				
Japanese encephalitis vaccine			X	Hoke CH et al, "Protection against Japanese encephalitis by inactivated vaccines", N Engl J Med. 1988 Sep 8;319(10):608-14.	Y	Y	Reduced encephalitis attack rate	Yes
Pneumococcal vaccine			X	Katherine L O'Brien et al, "Efficacy and safety of seven-valent conjugate pneumococcal vaccine in American Indian children: group randomised trial", Volume 362 Issue 9381 Page 355	Y	Y	prevents vaccine serotype invasive pneumococcal disease	Yes
Quadrivalent Vaccine			X	Efficacy of the Rhesus Rotavirus-Based Quadrivalent Vaccine in Infants and Young Children in Venezuela	Y	Y	Protection against severe diarrhea and dehydration, and reduced hospital admissions	Yes
Malaria			Y	Summarized in S. Meek, J. Hill, J. Webster, "The Evidence Base for Interventions to Reduce Malaria in Low and Middle-Income Countries", CMH Working Paper 6.	N	Y		
Indoor Spraying				Rowland M, et al. "Indoor residual spraying with alphacypermethrin controls malaria in Pakistan: a community-randomized trial", <i>Tropical Medicine and International Health</i> , 2000, Vol 5(7): 472-481.	Y	Y	Reduction in anopheline porous rates	Yes.
Insecticide-treated nets			X	Shulman CE, et al, "A community randomized controlled trial of insecticide-treated bednets for the preention of malaria and anaemia among primigravid women on the Kenyan coast" <i>Tropical Medicine and Internal Health</i> , 1998, 3 (3): 197-204	Y	Y	No significant impact	No. Not significant
Outdoor Spraying			X	David Cutler's slides	Y	N	Increase in arable land	No. Not random
HIV				Summarized in P. Jha, "The evidence base for interventions to prevent HIV infection in low and middle-income countries", CMH working paper WG5:2 (2001) and "Assessing the Economic Returns to Investing in Youths in Developing Countries" – James C. Knowles and Jere R. Behrman				
Voluntary HIV-1 counseling and testing			X	Coates et al, "Efficacy of voluntary HIV-1 counseling and testing in individuals and couples in Kenya, Tanzania, and Trinidad: a randomized trial". The Voluntary HIV-1 Counseling and Testing Efficacy Study Group. <i>Lancet</i> , 356:103-112 (2000)	Y	Y	Reduced intercourse with non-primary partners	Yes.
Condom provision in motel rooms			X	Egger M et al, "Promotion of condom use in a high-risk setting in Nicaragua: a randomized controlled trial". <i>Lancet</i> , 355:2101-2105 (2000)	Y	Y	Increased condom use	Yes

Package, including drug supply, health education and STD reference clinic			X	Grosskurth, H. et al, "Impact of improved treatment of sexually transmitted diseases on HIV infection in rural Tanzania: randomized controlled trial". <i>Lancet</i> , 346:530-536. (1995)	Y	Y	Reduced HIV incidence	Yes
Home-based mass antibiotic treatment			X	Wawer, M.J. et al, "Control of sexually transmitted diseases for AIDS prevention in Uganda: a randomized community trial." Rakai Project Study Group. <i>Lancet</i> , 353:525-535 (1999)	Y	Y	No effect.	No
Confidential HIV testing and condom promotion			X	Allen, S. et al, "Confidential HIV testing and condom promotion in Africa. Impact on HIV and gonorrhea rates." <i>JAMA</i> , 268:3338-3343.	Y	Y	Increased condom use, reduced rates of Gonorrhea and HIV in urban Rwandan women	Yes
Short-course zidovudine for babies that are not breastfed.			X	Shaffer N et al, "Short-course zidovudine for perinatal HIV-1 transmission in Bangkok, Thailand: a randomized controlled trial." <i>Lancet</i> , 1999:353:-773-780.	Y	Y	Reduced risk of mother-to-child HIV transmission	Yes
Short-course zidovudine for breastfed babies			X	Wiktor S.Z. et al, "Short-course oral zidovudine for prevention of mother-to-child transmission of HIV-1 in Abidjan, Côte d'Ivoire." <i>Lancet</i> , 1999, 353:781-785.	Y	Y	Reduced risk of mother-to-child HIV transmission	Yes
Monetary Transfers				Unlike the health-based interventions, all monetary transfers actually happened in a community-based context, hence they must be self-sustainable.				
Decentralized Targeting		X		Emanuela Galasso and Martin Ravallion, "Decentralized Targeting of an Anti-Poverty Program", forthcoming, <i>Journal of Public Economics</i> , Feb 2001	Y	N	Results better in more favorable conditions	No. Not program evaluation per se.
Old Age Pension			X	Esther Duflo, "Grandmothers and Granddaughters: Old Age Pension and Intra-household Allocation in South Africa", <i>World Bank Economic Review</i> , Vol 17(1), pp. 1-25, 2003	Y	N	Increased weight of girls	No. Not random
Poor-Area Development Programs		X		Jyotsna Jalan and Martin Ravallion, "Are There Dynamic Gains from a Poor-Area Development Program?", <i>Journal of Public Economics</i> , Jan 1998, 67(1), pp. 65-85	Y	N	Enough to prevent decline, but not enough for convergence	No. Not random.
Microfinance				Unlike the health-based interventions, all micro-finance interventions actually happened in a community-based context, hence they must be self-sustainable.				
Village Level Microfinance		X		Joseph Kaboski and Robert Townsend, "Policies and Impact: An Analysis of Village-Level Microfinance Institutions", University of Chicago, Oct 2002	Y	N	Mixed	No. Not random, not effective.
Village Level		X		Jonathan Morduch, "Does Microfinance Really Help	Y	N	No	No. Not

Microfinance				the Poor? New Evidence from Flagship Programs in Bangladesh”, Research Program in Development Studies, Princeton University, draft, 1998				effective.
School inputs				Unlike the health-based interventions, all school-based interventions actually happened in a community-based context, hence they must be self-sustainable.				
School Construction	X			Esther Duflo, “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment”, <i>American Economic Review</i> , Sept 2001	Y	N	Increase in education and earnings	No. Not random
School Construction	X			Esther Duflo, “The Medium Run Effects of Educational Expansion: Evidence from a Large School Construction Program in Indonesia”, forthcoming in <i>Journal of Development Economics</i> , Nov 2001	Y	N	Increase labor force participation , but reduce wages of older cohorts	No. Not random
Flip Charts	X			Paul Glewwe et al, “Retrospective vs. Prospective Analyses of School Inputs: The case of Flip Charts in Kenya”, NBER Working Paper 8018, Nov 2000	Y	N / Y	No evidence with prospective, positive with retrospective	No. No evidence using random evaluation
Uniforms	X			P. Glewwe, M. Kremer, S. Moulin, “Decentralization: A cautionary tale”, Harvard University mimeo, March 2003	Y	Y	Reduce dropout rates, without reducing test scores	Yes
Teacher incentives	X			Paul Glewwe, Nauman Ilias, M. Kremer, “Teacher Incentives”, Harvard University mimeo, 2003	Y	Y	No evidence, teachers teach to test	No. Not successful .
Textbook	X			Paul Glewwe, Michael Kremer, Sylvie Moulin, “Textbooks and Test scores: Evidence from a Prospective Evaluation in Kenya”, Harvard University mimeo, 2003	Y	Y	Raised test scores of the highest quintile, more likely to go to secondary school	No. Very skewed results.
Remedial Education Program	X			Abhijit Banerjee et. al, “Remedying Education: Evidence from Two Randomized Experiments in India”, MIT mimeo, July 2003	Y	Y	Increased learning	Yes.
Teacher provision	X			Aimee Chin, “The returns to school quality when school quality is very low: Evidence from Operation Blackboard in India”, mimeo, University of Houston, Feb 2002	Y	N	Increased female primary school completion and literacy	No. Not random.
School incentives								
School meals	X		X	Christel Vermeersch. “School Meals, Educational Achievement and School Competition: Evidence from a Randomized Experiment”, Harvard University mimeo	Y	Y	Conditional test score improvement, Higher	No. Questionable effectiveness

							participation , cut into instruction time	ess.
Progresa	X			Paul Schultz , “School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program”, Economic Growth Center, Center Discussion Paper No. 834, Aug 2001	Y	Y	Higher enrollments	Yes.
Progresa	X			Jere Behrman, Pilali Segupta and Petra Todd , “Progressing Through Progresa: An Impact Assessment of a School Subsidy Experiment”, Penn Institute for Economic Research, Working Paper 01-033, April 2001	Y	Y	Increased educational attainment, % attending junior secondary school	
Progresa			X	Jere Behrman and John Hoddinott, “Program Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican Progresa Impact on Child Nutrition”, Penn Institute for Economic Research, Working Paper 02-006, Nov 2001	Y	Y	0/- effect with means, + stature with Fixed Effects	
Progresa			X	P. Gertler, S. Boyce, “An Experiment in Incentive-Based Welfare: The Impact of PROGRESA on Health in Mexico”, Royal Economic Society Annual Conference 2003	Y	Y	Health improvements	
School vouchers	X			Joshua Angrist et al., “Vouchers for Private Schooling in Columbia: Evidence from an Randomized Natural Experiment” <i>American Economic Review</i> , Dec 2002, pp. 1535-1558	Y	Y	Higher completion, less repeating of grades	Yes.
Hygiene education	X		X	Haggerty, P.A et al. “Community-based hygiene education to reduce diarrhoeal disease in rural Zaire: impact of the intervention on diarrhoeal morbidity.” <i>International Journal of Epidemiology</i> 23(5):1050-1059. 1994	Y	Y	Reduced diarrhea	Yes. (Folded into other education Estimates)
Fertilizer								
Fertilizer adoption through an NGO program				Esther Duflo and Michael Kremer, “Understanding Technological Choices: Fertilizers in Western Kenya”, power-point slides	Y	Y	Increased adoption, high rate of return	Yes
Roads								
Rural Road Construction				Dominique van de Walle, “Impact Evaluation of a Rural Road Rehabilitation Project in Viet Nam”, http://econ.worldbank.org/view.php?type=20&id=11865	Y	N	Research in progress.	Not random, research in progress

Notes

E: Education
P: direct Poverty reduction
H: Health
Pr: Program sustainable?
R: Random?

-
- ⁱ Year of 1st independent evaluation as given by Jan 2003 “Catalogue of DFID Evaluation Studies”
- ⁱⁱ Para 118, Michael Flint et al, “How effective is DFID? : Development effectiveness report 2001”, 2nd Draft, DRID, March 30th, 2002.
- ⁱⁱⁱ Para 115, Michael Flint et al, “How effective is DFID? : Development effectiveness report 2001”, 2nd Draft, DRID, March 30th, 2002.
- ^{iv} Para 116, Michael Flint et al, “How effective is DFID? : Development effectiveness report 2001”, 2nd Draft, DRID, March 30th, 2002.
- ^v “Agency Reorganization: Generic Functional Statements.”
<http://www.usaid.gov/about/reform/functions.html>
- ^{vi} Cynthia Clapp-Wincek, Richard Blue, “Evaluation of Recent USAID Evaluation Experience”, Center for Development Information and Evaluation, Working paper No. 320, page iii.
- ^{vii} Cynthia Clapp-Wincek, Richard Blue, “Evaluation of Recent USAID Evaluation Experience”, Center for Development Information and Evaluation, Working paper No. 320, page iii.
- ^{viii} From Cynthia Clapp-Wincek, Richard Blue, “Evaluation of Recent USAID Evaluation Experience”, Center for Development Information and Evaluation, Working paper No. 320, page 37. Here we face the related problems of
- (a) Few evaluations done: “Of most concern is the very limited number of in-depth, program evaluations”
- (b) Lack of learning even if evaluations are carried out: “Most of the evaluation work that is being done is being done by partners. The partner organizations are learning from the experience; USAID is not”.
- (c) Lack of independence: The USAID managers decide if they want to do the evaluations.
- (d) The Fly-in approach: “Scopes ask a team to come for 4-6 weeks and interview the mission, the activity staff, and ‘representatives’ of the local people. There isn’t enough time to get any kind of representative sample. The team frequently tells the USAID manager pretty much what he already knows”
- ^{ix} The ADS 200 series (Sept. 2000) added a new dimension to evaluations. The “Reform Vision” in ADS 200 states these expectations: “Applying the lessons of successes and failure systematically and providing leadership in tackling complex problems that demand multi-agency or multi-donor responses.” (Quoted in Cynthia Clapp and Richard Blue (see previous footnote), page 1
- ^x Jean DuRette, Glenn Slocum, “The Role Of Transitional Assistance: The Case Of East Timor”, U.S. Agency for International Development, November 2001.
- ^{xi} “FY 2000 Performance Overview” United States Agency for International Development. See Page 29 for example.
- ^{xii} “USAID Performance and Monitoring Report, FY 2002”. http://www.dec.org/pdf_docs/PDABY087.pdf Appendix 2.
- ^{xiii} http://www.afdb.org/knowledge/documents/ADB_in_brief.htm
- ^{xiv} “African Development Bank Group”, http://www.ustreas.gov/offices/international-affairs/intl/fy2003/tab10_afdbg_afdb.pdf, United States Department of the Treasury.
- ^{xv} African Development Bank Group “Strategic Plan 2003-2007”,
http://www.afdb.org/knowledge/publications/pdf/adb_strategic_plan2003-2007e.pdf
 November 2002, page 52
- ^{xvi} <http://www.dac-evaluations-cad.org/dac/>
- ^{xvii} Asian Development Bank, Operations Evaluation Department. “Annual review of evaluation activities in 2002”, page 61.
- ^{xviii} Frequently Asked Questions. <http://www.adb.org/Evaluation/faqs.asp>
- ^{xix} All reports online since 1995. <http://www.adb.org/Evaluation/reports.asp>
- ^{xx} “Annual Review of Evaluation Activities in 2002”, Asian Development Bank, Operations Evaluation Department, May 2003, Page 67.
- ^{xxi} “Annual Review of Evaluation Activities in 2002”, Asian Development Bank, Operations Evaluation Department, May 2003, Page 47.
- ^{xxii} Asian Development Bank, Operations Evaluation Department. “Annual review of evaluation activities in 2002”, page 47.
- ^{xxiii} PPMS Project Framework and Performance Indicators 2.
http://www.adb.org/Documents/Slideshows/PPMS/4b_PPMS_Indicators.pdf
- ^{xxiv} Project Evaluation Department Brouchure, August 1999, European Bank for Reconstruction and Development, page 6.

-
- ^{xxv} Project Evaluation Department Brouchure, August 1999, European Bank for Reconstruction and Development, page 6.
- ^{xxvi} Annual Report 2002: Annual Review and Financial Report, European Bank for Reconstruction and Development, page 70.
- ^{xxvii} Private Communication, Fredrik Korfker, Corporate Director, Evaluation, EBRD. Sept 9 2003.
- ^{xxviii} Project Evaluation Department Brouchure, August 1999, European Bank for Reconstruction and Development, page 12.
- ^{xxix} Project Evaluation Department Brouchure, August 1999, European Bank for Reconstruction and Development, page 12.
- ^{xxx} Evaluation Office, OVE, “Annual Report of the Evaluation Office 2000”, Inter-American Development Bank, Washington D.C., June 2001, RE-249, Chapter 2.1
- ^{xxxi} <http://www.iadb.org/ove/re268eng.pdf>, page 28
- ^{xxxii} <http://www.iadb.org/ove/re259eng.pdf>, page 29
- ^{xxxiii} <http://www.iadb.org/cont/evo/ovedocs.htm>
- ^{xxxiv} Evaluation Office, OVE, “Annual Report of the Evaluation Office 2000”, Inter-American Development Bank, Washington D.C., June 2001, RE-249, Chapter 2.1
- ^{xxxv} <http://www.iadb.org/ove/re262eng.pdf> page 13
- ^{xxxvi} Office of Evaluation and Oversight, OVE “Country Program Evaluation, Costa Rica: 1990-2001” <http://www.iadb.org/ove/re277eng.pdf> See footnote 20.
- ^{xxxvii} Office of Evaluation and Oversight, OVE, “Annual Report of the Office of Evaluation and Oversight 2001”, Inter-American Development Bank, Washington D.C., November 2002. <http://www.iadb.org/ove/re268eng.pdf>, page 7.
- ^{xxxviii} Office of Evaluation and Oversight, OVE “Country Program Evaluation, Costa Rica: 1990-2001” <http://www.iadb.org/ove/re277eng.pdf> See footnote 20.
- ^{xxxix} Office of Evaluation and Oversight, OVE “Country Program Evaluation, Costa Rica: 1990-2001” <http://www.iadb.org/ove/re277eng.pdf> See page 40. Out of the 4 categories of projects, only the Public Reform Sector has PCRs.
- ^{xl} Office of Evaluation and Oversight, OVE, “Annual Report of the Office of Evaluation and Oversight 2001”, Inter-American Development Bank, Washington D.C., November 2002. <http://www.iadb.org/ove/re268eng.pdf>, page 7.
- ^{xli} Evaluation Group of Executive Directors, “Review of Experience with Evaluation in the Fund”, <http://www.imf.org/external/np/eval/2000/031400.HTM>, March 14, 2000
- ^{xlii} Evaluation Group of Executive Directors, “Review of Experience with Evaluation in the Fund”, <http://www.imf.org/external/np/eval/2000/031400.HTM>
- ^{xliii} Unlike the other agencies, the IMF looks after Macro-economic performance, hence country-level evaluations make sense here.
- ^{xliv} Oct 31, 2001. “Progress in making the independent evaluation office (IEO) operational.” <http://www.imf.org/external/np/eval/2001/103101.htm>
- ^{xlv} Sept 12, 2000. “IMF Executive Board Report to the IMFC on the Establishment of the Independent Evaluation Office (EVO) and its Terms of Reference” <http://www.imf.org/external/np/eval/2000/091200.htm>
- ^{xlvi} Independent Evaluation Office, International Monetary Fund: “Fiscal Adjustment in IMF-Supported Programs”, Sept 2003, “IMF and the Recent Capital Account Crises: Indonesia, Korea, Brazil”, July 2003 and “Evaluation of Prolonged Use of IMF Resources”, 2002
- ^{xlvii} “IEO Annual Report 2003”, IMF, Sept 2003, Page 10. <http://www.imf.org/external/np/ieo/2003/ar/Report.pdf>
- ^{xlviii} “Evaluation: Report of the Administrator”, Executive Board of the United Nations Development Programme and of the United Nations Population Fund. 2000. <http://www.undp.org/execbrd/pdf/dp00-34e.pdf> page 21.
- ^{xlix} “Country offices have the lead role in determining which projects at country level are to be evaluated; however, all projects with budgets of more than \$1 million must be evaluated”. Page 15, UNDP Evaluation Office, “Development effectiveness: Review of evaluative evidence”.

-
- ⁱ Evaluation: Report of the Administrator”, Executive Board of the United Nations Development Programme and of the United Nations Population Fund. 2001 . <http://www.undp.org/execbrd/pdf/dp01-26e.pdf> Table 6
- ⁱⁱ “Handbook on monitoring and evaluating for results.” United Nations Development Program, Evaluation Office, Part 1, Table 3.
- ⁱⁱⁱ Handbook on monitoring and evaluating for results.” United Nations Development Program, Evaluation Office, Part 2, Page 45.
- ^{liii} “Development Effectiveness, Review of Evaluative Evidence” United Nations Development Office Evaluation Office, November 2001, Page 15.
- ^{liv} United Nations Development Program Evaluations Database.
<http://www.undp.org/eo/database/index.htm>
- ^{lv} “Development Effectiveness, Review of Evaluative Evidence” United Nations Development Office Evaluation Office, November 2001, Page 15 to 19.
- ^{lvi} “Development Effectiveness, Review of Evaluative Evidence” United Nations Development Office Evaluation Office, November 2001, Page 35.
- ^{lvii} “Development Effectiveness, Review of Evaluative Evidence” United Nations Development Office Evaluation Office, November 2001. Page 19.
- ^{lviii} “Development Effectiveness, Review of Evaluative Evidence” United Nations Development Office Evaluation Office, November 2001. Page 15.
- ^{lix} “Development Effectiveness, Review of Evaluative Evidence” United Nations Development Office Evaluation Office, September 2000. Page 10.
- ^{lx} “OED: The first 30 years”, The World Bank, Operations Evaluation Department,
http://www.worldbank.org/wbi/B-SPAN/docs/oed_thirty.pdf
- ^{lxi} “2001 Annual Review of Development Effectiveness”, Operations Evaluations Department, World Bank, May 2002, Annex D.
- ^{lxii}
<http://lnweb18.worldbank.org/servlet/OEDSearchServlet?SearchType=byField&PerPage=20&DbURL=oe/oeddolib.nsf&Series=OED%20Project%20Evaluation>
- ^{lxiii} “2002 Annual Review of Development Effectiveness”, Operations Evaluations Department, World Bank, Feb 2003, Table 5.
- ^{lxiv} “2002 Annual Review of Development Effectiveness”, Operations Evaluations Department, World Bank, Feb 2003, Table 14.
- ^{lxv} “2002 Annual Review of Development Effectiveness”, Operations Evaluations Department, World Bank, Feb 2003, Table 14.
- ^{lxvi} “2002 Annual Review of Development Effectiveness”, Operations Evaluations Department, World Bank, Feb 2003, Table 14.