

# *POST-WASHINGTON CONSENSUS TASK FORCE REPORT*

## *DOES DEVELOPMENT AID MAKE A DIFFERENCE?*

*ESTHER DUFLO*

*NOVEMBER 30, 2003*

“Does Development Aid Make a Difference?” Whether we are in the business of giving aid, raising funds, or are simply well intentioned observers of a world where so many seem in dire need of assistance, this is a question we would really like to have an answer to. Not surprisingly then, many have tried to answer it, usually after specifying it as “Does development aid promote growth?” While the answer varies from author to author the tone that prevails these days seem to be pessimistic. Even Burnside and Dollar’s (2000) carefully optimistic conclusion: that aid raises growth in countries with good policy environments, but not in countries with bad policy environments, has come under attack from a variety of fronts. Most recently by Easterly, Levine and Roodman (2003) who run Burnside and Dollar’s regressions using an expanded data set, and find that the coefficient of aid turns negative (though insignificant) even in countries with a good policy environment. More generally, many observers such as Easterly (2001, 2003) lament the ineffectiveness of aid and point to factors ranging from the lack of incentives for donors to monitor aid effectiveness, to the way aid bureaucracy is organized and interacts with the aid recipients. Yet there is no doubt that some aid, sometimes makes an enormous difference. The initial investments into the green revolution have generated returns that are probably high enough to pay for all the foreign aid that has been disbursed since then. Some NGO-run programs, such as de-worming in Kenya have been shown to have tremendous rates of returns. In this paper, I argue that looking for the impact of aid on growth is probably a red herring; that, if one is to find out whether aid makes a difference, one has to go back to the more traditional public finance cost-benefits calculations; that to make these cost benefits calculations in an informed way, one needs credible estimates of the impact of a wide variety of projects, combined with an assessment of the effectiveness of delivery for each project; and that for many programs targeted to individuals or communities, randomized evaluation is a feasible, cost-effective way to generate credible estimates of project impact. In conclusion, I argue that spending more energy and resources to identify whether, project by project, development aid makes a difference could have two major effects - valuable over and above their role as an input in calibration of overall aid effectiveness. First, it would improve effectiveness, by providing information on the basis of which future program allocations can be made, and second it would dispel some of uncertainty and skepticism that surrounds aid agencies, and reassure donors, thus increasing the overall aid budget. Projects repeatedly shown to be ineffective could be discontinued, and effective projects could be expanded.

### **1. Is “does development aid increase growth” a good substitute for “does development aid make a difference?”**

At first pass, the question of whether development aid makes a difference must be obvious. It typically involves redistribution from richer countries to poorer countries, and in this sense, it is likely to improve aggregate welfare as long as the money does not immediately go back to Switzerland. However, we typically feel that simply increasing consumption is not the goal of development aid: aid must translate into productive investments in the receiving country. The following quote by Burnside and Dollar (2003) reflects this general sentiment “One interpretation of foreign aid is that it is an income transfer. This

transfer may or may not produce growth. Is it invested, so that domestic product can increase, or is it consumed? To the extent that it is invested, aid will be effective.” The best way to assess whether a country has engaged in productive investments is to then measure its growth. This has the advantage of being a summary measure for the country, which can be coupled with the overall amount of aid available. Finding out the effectiveness of each separate project becomes thus unnecessary, since in the neo-classical model, growth is an adequate summary statistic of all investments that took place in the economy.

Looking for the impact of aid only through the lens of its effect on growth presents several conceptual and practical problems, however. Conceptually, the distinction between consumption and investment may be blurred. It is estimated that 6 million of the 10.8 million child deaths that occur every year are preventable with cheap, low-tech methods. The World Bank estimates the cost-per-year of a life saved using health interventions such as vaccination against measles, oral rehydration therapy, or bed nets, range from \$5 to \$40. Using the numbers commonly employed to value a year of life in the United States, the rate of return to these investments would be enormous. Yet saving children may not increase growth per capita, and may even reduce it, if the increase in child survival is not matched by a decrease in fertility. Does it follow that any aid spent on these technologies makes no difference? Timing is another tricky issue. In Burnside and Dollar (2003), for example, aid and growth are both averaged over a contemporaneous four-year period. Conceivably, any effect on growth would take much longer to be felt.

These are just two examples of the conceptual difficulties, but the practical problems are, in my view, more important. The main practical problem (of some relevance) is that the question “does development aid increase growth” cannot really be answered. To use economists’ jargon, it is not identified; that is, it is very unlikely that someone will come up with a convincing estimate of the causal effect of aid on growth. The reason is evident, and clearly stated by Burnside and Dollar (2003): aid may be a function of growth in the country. This will induce a tendency for aid to be associated with growth, “to the extent that donors respond to negative growth shocks by providing more assistance. But there are plausible reasons why the errors may have a positive correlation. One conclusion of earlier studies and our own work is that aid is not given only for developmental purposes; it may serve the strategic or commercial interests of donors. In that case a country enjoying a commodity boom, or any positive shock to growth, may receive special favor from some donors, introducing a positive correlation between the error terms of the period.” This means that the statement that, “on average countries grow more (or less) slowly if they get more aid,” does not answer the question of interest. In the same way that it makes no sense for the World Bank to claim that its aid is effective because China grew rapidly (World Bank, 2002), it also makes no sense to say that aid is ineffective because aid to Africa has steadily increased between 1970 and 1996 while growth plummeted. Aware of this problem, some authors have tried to examine the correlation between growth and the part of aid that can be predicted by variables that do not directly predict growth.

But where can we find such variables? Burnside and Dollar (2003) use a country’s population, arms imports, being Egypt, belonging to the “zone franc” and being part of Central America to predict aid, and argue these variables do not predict growth directly. Unfortunately, there is no legitimate argument to exclude any of these variables *a priori* from determining growth over and above their impact on aid. For example, what guarantees that Egypt’s growth would have been exactly average if it did not receive more aid than others; Shleifer and co-authors have repeatedly argued that ex-French colonies have worse growth performance than others, not because of aid, but because of civil law; why should we not expect that arms trafficking has something to do with a country’s growth; what about population? While I cannot rule out *a priori* that it will ever be possible to credibly isolate components of aid linked to variables that do not directly predict growth, I must say that I am pessimistic. Until then, back and forth arguments about whether aid predicts growth will remain somewhat empty.

## 2. Does development aid make a difference? How to find out?

How then, to find out whether development aid makes a difference? The arguments against using growth as a benchmark essentially apply to the most common alternatives, looking for an impact of aid on the poverty rate, or the income of the poor in the countries that receive aid. The problem is compounded in this case by the fact that measuring poverty is far from evident, and these exercises are fraught with their own controversies. See Deaton (2003) for a nice summary.

In the absence of convenient summary statistics, one has to go back to a traditional cost-benefits analysis of the impact of each project (perhaps after weighting the benefits according to distributional objectives), and aggregate the rates of returns to obtain an estimate of the effectiveness of the overall aid provided. Unfortunately, as Banerjee and He (2003) show, most large donor organizations do not provide quantitative assessments of the social impact of their projects, such as a rate of return. Of the major donors surveyed by Banerjee and He, only the World Bank reports rates of return and then only for certain sectors. Education, health and nutrition, for example, are left out. Only the World Bank and the Asian Development Bank have their projects assessed formally by an independent organization, report the results on a uniform scale, and allow public access to their results. Aid organizations' reluctance to evaluate their programs not only reflects the difficulty in measuring a program's impact, and to aggregate its benefits on a set scale, but also underlines the lack of incentives to objectively assess program effectiveness. This is particularly true for funding organizations that are perpetually in search of new funds and would rather avoid bad publicity (Pritchett, 2002). The lack of pressure by ultimate donors (e.g. government) for credible evaluations should not be ignored, and perhaps be attributed to the difficulty in discriminating between good and bad evaluations (Kremer 2003).

How should one proceed to construct credible estimates of aid effectiveness? The impact of some macroeconomic measures (debt forgiveness or technical assistance to the central bank, for example) may not be easily estimated, since it would be very difficult to compare a given country with or without this aid. But a substantial fraction of aid is given in the form of programs that target people or communities. Even general budget support, currently advocated by many organizations, typically results in a series of programs being undertaken by the government. Most of the programs implemented by international institutions, NGOs, or governments, are not currently evaluated, or if they are, are evaluated in ways that do not allow for their true effectiveness to be measured. Many programs are subjected to a *process evaluation* (did the program actually happen on the ground?). Some are also subjected to an *impact evaluation* (did the program actually affect the final outcome?), but for reasons we develop below, these impact evaluations are often not credible. Not every single program can or should be subjected to an *impact evaluation* (in particular when several previous impact evaluations have convincingly shown its results), but to measure the effectiveness of the share of aid that ultimately goes to individual projects, I propose to combine a process evaluation for each project with our best estimate of the proposed impact of the program in this context, deriving either from a rigorous program evaluation of that particular program, or using evidence from similar programs that have been evaluated in the past.

## 3. The Evaluation Problem

Any impact evaluation attempts to answer an essentially counterfactual question: how would individuals who participated in the program have fared in the absence of the program? How would those who were not exposed to the program have fared in the presence of the program? The difficulty with these questions is immediately apparent: at a given point in time, an individual is observed to be either exposed or not exposed to the program. Comparing the same individual over time will not, in most cases, give a

reliable estimate of the impact the program had on him or her, since many other things may have changed at the same time as the program was introduced. We cannot therefore seek to obtain an estimate of the impact of the program on each individual. All we can hope for is to be able to obtain the average impact of the program on a group of individuals by comparing them to a similar group of individuals who were not exposed to the program.

The critical objective of impact evaluation is therefore to establish a credible comparison group, a group of individuals who in the absence of the program would have had outcomes similar to those who were exposed to the program. This group should give us an idea of what would have happened to the members of the program group if they had not been exposed, and thus allow us to obtain an estimate of the average impact on the group in question.

In reality, however, the individuals who participated in a program generally differ from those who did not: programs are placed in specific areas (for example, poorer or richer areas), individuals are screened for participation in the program (for example, on the basis of poverty or on the basis of their motivation), and, in addition, the decision to participate is often voluntary. For all of these reasons, those who were not exposed to a program are often a poor comparison group for those who were, and any differences between the groups can be attributed to two factors: pre-existing differences (the so-called “selection bias”) and the impact of the program. Since we have no reliable way to estimate the size of the selection bias, we typically cannot decompose the overall difference into a treatment effect and a bias term.

To solve this problem, program evaluations typically need to be carefully planned in advance in order to determine which group is a likely control group. One situation where the selection bias disappears is when the treatment and comparison groups are selected randomly from a potential population of participants (such as individuals, communities, schools, or classrooms). In this case, on average, we can be assured that those who are exposed to the program are no different than those who are not, and thus that a statistically significant difference between the groups in the outcomes the program was planning to affect can be confidently attributed to the program.

#### **4. The use of Randomization in the Evaluation of Development Effectiveness**

At the moment, randomized evaluations are only used to evaluate a tiny fraction of development programs. If we apply the method I proposed earlier, of combining a process evaluation of a development program with an impact evaluation of similar programs, we should be able to assess the effectiveness of a very small fraction of overall aid disbursed. This share would be an interesting number to compute. It would make clear how little we know about aid effectiveness, and what fraction of aid is thus spent by shooting in the dark. Yet, Banerjee and He (2003) calculate that, even if we scaled up only the programs for which we have evidence that they are effective, based on randomized evaluation, it would be possible to spend at least 11 billion dollars (and this does not include PROGRESA which they treat as an income transfer).

There may be even better use of this money, but at the moment we just do not know. Elsewhere (Duflo (2003), Duflo and Kremer (2003)), I have argued that randomized evaluations are often feasible, do not require larger budget than other kinds of evaluations, and provide different (and better) answers than other kinds of evaluation. International institutions should have a key role in promoting them. Rigorous and systematic evaluations have the potential to leverage the impact of international organizations well beyond simply their ability to finance programs. Credible impact evaluations are international public goods: the benefit of knowing that a program works or does not work extends well beyond the organization or the country implementing the program. Programs that have been shown to be successful can be adapted for use in other countries and scaled-up within countries, while unsuccessful programs

can be abandoned. Through promoting, encouraging, and financing rigorous evaluations (such as credible randomized evaluations) of the programs they support, as well as of programs supported by others, international organizations can provide guidance to the international organizations themselves, as well as other donors, governments, and NGOs in the ongoing search for successful programs. Moreover, by credibly establishing which programs work and which do not, international agencies can counteract skepticism about the possibility of spending aid effectively and build long-term support for development. Just as randomized trials revolutionized medicine in the 20th century, they have the potential to revolutionize social policy during the 21st.

## REFERENCES

Banerjee, Abhijit and Ruimin He (2003) "Making aid work", Mimeo, MIT, October

Burnside, Craig and Dollar, David. "Aid, Policies, and Growth (2000)" *American Economic Review*, September, 90(4), pp. 847–68.

Deaton, Angus (2003) "Measuring Poverty in a Growing World" (or "Measuring Growth in a Poor World") NBER Working Paper 9822

Duflo, Esther (2003), "Scaling Up and Evaluation," Annual World Bank Conference in Development Economics Conference Proceedings. Washington DC: World Bank, forthcoming.

Duflo, Esther and Michael Kremer (2003) "Use of Randomization in the Evaluation of Development Effectiveness," forthcoming, *Proceedings of the Conference on Evaluating Development Effectiveness, July 15- 16, 2003*, the World Bank Operations Evaluation Department (OED) Washington, D.C.

Easterly, William (2001) *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*. Cambridge MA: MIT Press, 2001

Easterly, William "Can Foreign Aid Buy Growth?", *Journal of Economic Perspectives*, forthcoming, 2003

Easterly, William, Ross Levine and David Roodman (2003) (with R. Levine and D. Roodman) "New data, new doubts: A Comment on Burnside and Dollar's "Aid, Policies, and Growth" (2000), forthcoming *American Economic Review*, 2003

World Bank (2002) World Bank (2002) *The Role and Effectiveness of Development Assistance: Lessons from World Bank Experience*. Washington D.C.: World Bank.