Stockholm Institute of Transition Economics

WORKING PAPER

June 2012

No. 16

The Politics of Aid Effectiveness: Why Better Tools can Make for Worse Outcomes

Anders Olofsgård

The Politics of Aid Effectiveness: Why Better Tools can Make for Worse Outcomes*

Anders Olofsgård†
Georgetown University

Preliminary. This version: June 26, 2012

Abstract

The recent focus on impact evaluation within development economics has lead to increased pressure on aid agencies to provide "hard evidence", i.e. results from randomized controlled trials (RCTs), to motivate how they spend their money. In this paper I argue that even though RCTs can help us better understand if some interventions work or not, it can also reinforce an existing bias towards focusing on what generates quick, immediately verifiable and media-packaged results, at the expense of more long term and complex processes of learning and institutional development. This bias comes from a combination of public ignorance, simplistic media coverage and the temptation of politicians to play to the simplistic to gain political points and mitigate the risks of bad publicity. I formalize this idea in a simple principal-agent model with a government and an aid agency. The agency has two instruments to improve immediately verifiable outcomes; choose to spend more of the resources on operations rather than learning or select better projects/programs. I first show that if the government cares about long term development, then incentives will be moderated not to push the agency to neglect learning. If the government is impatient, though, then the optimal contract leads to stronger incentives, positively affecting the quality of projects/programs but also negatively affecting the allocation of resources across operations and learning. Finally, I show that in the presence of an impatient government, then the introduction of a better instrument for impact evaluation, such as RCTs, may actually decrease aid effectiveness by motivating the government to chose even stronger incentives. (*Keywords: Foreign aid, principal agent, political economy*)

*I am grateful for comments and suggestions from Maria Perrotta and seminar participants at Georgetown University and the Stockholm School of Economics.

†Associate Professor, Georgetown University and Deputy Director, SITE, Stockholm School of Economics. afo2@georgetown.edu

1 Introduction

In implementing aid policy there is a risk of conflict between immediately verifiable outcomes and more long term effects from learning and institutional development. For the immediate effects, the effectiveness and the role of aid is typically easier to quantify with reasonable accuracy. For the public goods of learning and institutional development, the impact is typically more long run, harder to quantify and harder to directly attribute to the inflow of aid. A conflict arises when limited time and financial resources force practitioners to trade off activities that maximize the chances of immediate benefits against those that may yield more long term effects. For instance, immediate results may be more likely if financial resources are spent on foreign technical assistance, whereas the training of partner country personnel may foster more learning and institutional development. In some projects learning is the explicit focus, but in most cases this is more a by-product, and its extent depends on how much time and effort is devoted to incorporating local partners and institutions, and making sure that the project is sustainable also when aid flows cease. In other words, it doesn't only matter what is done, but also how it is done. To maximize the long run developmental impact of aid, the allocation of resources within projects and programs between what I will henceforth refer to as operations and learning respectively (for lack of better expressions) needs to strike a fine balance.

That striking such a balance is not a trivial task is immediately apparent when tuning in to the debate on the history and future of foreign aid.[1] Arguments about what reforms should be made, and who is to blame for past failures, are almost as many as there are authors of books and articles on the subject. Of particular relevance for this paper are two lines of arguments focusing on impact evaluation and institutional development. The first line of argument is that aid agencies, claimed

[1] See for instance Sachs 2005 and Easterly 2006 for two detailed but very different accounts of aid's past effectiveness and future potential.

to be suffering from institutional inertia, weak incentives and too much focus on internal processes rather than on external impact, are slow to adjust practices to what research shows generates directly measurable positive effects. This argument is fueled by the current focus on randomized controlled trials (RCTs) within micro oriented development research. The rapid expansion of RCTs of social programs has created a bank of alledgedly internally valid knowledge of the immediate impact of different interventions in different sectors, such as education, health and financial markets. This knowledge should be used to focus aid on projects with proven track records, the argument goes, and the responsibility for implementation, quality and evaluation should primarily be in the hands of donor agencies.[2]

Another line of debate, with some foundations in the more macro oriented development literature, argues that the lack of patience with long run institutional development is the major culprit for why aid often fails to generate desirable results. This impatience in turn is blamed on the complexity in evaluating the long run impact of aid together with political pressure to show more or less immediate results to please domestic constituencies. Proponents of this view typically emphasize the need for recipient country ownership, and argue that aid should be channeled more as program aid, through partner country financial systems, and with more emphasis on learning and domestic responsibility. This does not preclude that aid finances particular projects, but in that process knowledge spillovers and institutional improvements that are harder to evaluate may be just as important as more easily measured outcomes. These arguments are part of the background behind the attempts to strengthen ownership and make donors use local systems through the Paris Declaration and the Accra Agenda.

In this paper I use a simple principal-agent model to illustrate how the political ramifications

[2]This view is for instance captured in the following quote: "From this point of view the current fashion for channeling aid into broad budgetary support (rather than specific projects) in the name of national autonomy seems particularly disastrous. We need to go back to financing projects and insist that the results be measured." (Banerjee, 2007, p. 21).

of aid are crucial for how to thread the line between immediate impact and learning. The model assumes resistance to change within the agency, so an incentive contract is necessary to motivate the agency to accommodate its project/program platform in response to new evidence of what interventions have the highest expected rate of return. Better immediate outcomes can however also be achieved by a reallocation of resources from learning to operations, i.e. a change in *how* the projects/programs are implemented rather than *which* projects/programs are financed. If the government (principal) does not have the resources to judge to what extent a change in immediate output comes from a change in the types of projects/programs or from a change in how the projects/programs are implemented, then the incentive contract needs to consider this private information. If the government is patient and cares about learning and institutional development in addition to immediately verifiable output, then incentives will be moderated to avoid excessive focus on operations. The level of moderation will depend on the principal's ability to determine to what extent the outcome depends on the actions of the agency rather than by chance, the standard trade off between incentives and risk. The introduction of a new evaluation technology, such as RCTs, will lead to stronger incentives since the agency is more receptive to incentives when there is a tighter match between effort and output. In the case of a patient government, the improved evaluation technology will unambiguously improve the development outcome by permitting stronger incentives, while still internalizing the impact on learning and institutional development. On the other hand, if the government is impatient and only cares about immediately observable outcomes then the effect on learning is not internalized. It follows that the contract will have stronger incentives, motivating more effort to implement better projects/programs but also a deterioration in learning and institutional development. Most importantly, in this case a better evaluation technology will have two counteracting forces. On the one hand it will motivate better selection of

projects/programs, but, on the other hand, with stronger incentives in the contract, the negative externality on learning and institutional development will also be more severe. If the latter effect dominates, improved evaluation techniques may even decrease the developmental impact of aid by causing an excessive emphasis on the short run impact.

What the simple model illustrates is that in a second best world where politicians are too impatient, or under too much political pressure to generate immediately verifiable results, the introduction of a technology to more accurately evaluate the developmental impact of aid projects may actually be counterproductive. An innovation that would normally improve effectiveness of aid instead reinforces an existing pattern of misallocation of resources. This result points to the importance of considering the political and institutional reality when offering policy recommendations based on academic advancements. The first best option is of course to have patient politicians and use the better evaluation technology, but if the former is not realistic, then one needs to be cautious in arguing that the possibility to randomize should be a requirement for financing until also learning and institutional development can be measured with the same accuracy.

The reminder of the paper is organized as follows. Section two brings up more in detail two current themes in the debate on aid effectiveness and the role of RCTs. In section three the simple model is developed and solved. Section four concludes and discusses the implications of the results.

2 Aid Effectiveness and Aid Evaluation

The purpose of this section is to discuss more in detail the two lines of debate brought up in the Introduction. In particular, I will argue that the two different arguments not only differ in terms of where they think the focus of aid should be, but also in terms of what they identify as the reason for current shortcomings, and who is responsible. This means that the suggested solutions differ

not only in terms of how resources should be reallocated across tasks, but also in terms of how to allocate authority across agencies and politicians, and to what extent RCTs is the long sought for panacea.

2.1 Narratives of Aid Effectiveness

According to the first narrative, a main problem with current aid policy in most western donor countries is that aid agencies are reluctant to change, overly bureaucratic and slow to adopt their strategies in response to new developments from within the academic community. Agencies are accused of institutional inertia; programs and projects keep getting financed despite doubts about their effectiveness because agency staff and aid contractors are financially and emotionally attached. This inertia is made possible by a lack of accountability and transparency towards tax payers as well as the ultimate beneficiaries. These shortcomings partly come from the special nature of aid as such. For instance, contrary to the case of most public goods, the beneficiaries of aid cannot hold the providers politically accountable since they reside in different countries (Seabright 2002). Also, aid fungibility, crowding-out effects, multiplicity of objectives, and the fractionalization of aid across donors and projects are all factors that complicate a proper evaluation of the extent to which aid actually contributes to long run development. However, even if part of these problems may be inherent to the nature of aid as such, parts of it certainly depends on how aid is organized, and how aid agencies operate. Both Easterly and Pfutze, 2008, and Birdsall and Kharas, 2010, report a lack of response from most aid agencies when asked for information pertaining to different aspects of how they spend their money. Only 10 out of 31 agencies pass Easterly and Pfutze's "transparency test".[3]

[3] In response to what they perceived as lack of transparency in the handling of aid policies within Sida, the new Swedish government launched an initiative they refer to as "open aid" in 2010. Included is a "transparency guarantee" which states that it should be possible for everybody to get information about what is done, and why. In particular, all information that is not confidential should be available on internet.

A reason for why agencies prefer less transparency and accountability, according to the critics, is that it yields the agencies more discretion in how to spend the money. Reluctance from within aid agencies is interpreted as a sign of "institutional laziness", or a concern that revealing the true impact of aid would decrease the support for aid among politicians and the general public. In this narrative, aid effectiveness would benefit from stronger incentives for aid agencies to generate verifiable outcomes, and politicians and tax payers should be given better instruments to hold agencies accountable.

In the second narrative, a main problem with aid policy is the impatience with institution building (e.g. Birdsall, 2004). Development is seen as a long term process of creating and sustaining strong economic and political institutions, not just individually successful aid financed social projects. The contributions of foreign aid also come in forms of learning, cultural immersion, and changing norms, outcomes that are just as important but much harder to evaluate. Parts of this is explicit as in technical assistance for institution building, but much of it is implicit in the way projects are implemented, with local participation and ownership being crucial for the build-up of experience, know-how and a sense of responsibility. The importance of institutions for long run development is the leading theme within more macro-oriented development research (e.g. Acemoglu et al. 2001 and Rodrik et al. 2004), so the emphasis on institutions has scholarly support, even though the link from aid to better institutions has less foundations.

Ownership and learning can go counter to what maximizes the chances of a successful project in the short run, though. This is shown by the persistence of donors to create so called project implementation units (PIUs), basically parallel development management structures outside recipient country government agencies. These structures contribute to the well known problem that aid projects often draw scarce domestic resources from more urgent tasks in the development process,

in particular skilled personnel eroding the capacity of recipient governments (Knack and Rahman, 2007).[4] Aware of this problem, donors committed in the Paris Declaration to reduce the use of PIUs, but the view seems to be that progress has been slow so far (e.g. Birdsall and Kharas, 2010).

So where does this impatience with institution building come from? The typical explanation is that it comes from a political need to show immediate results. This causes a bias towards actions that first of all generate direct results, and secondly that can be relatively easily evaluated. This creates the risk of bias towards project rather than program aid, and evaluations tend to focus on a narrow set of easily quantifiable outcomes rather than the harder to measure and more long run effects on deeper economic and social development. Knack and Rahman (2007, p. 2) argue that aid agencies, in order to satisfy domestic constituencies in parliament, are pushed into "..making the results of aid programs visible, quantifiable, and directly attributable to the donor's activities - even when doing so reduces the developmental impact of aid". A line minister in charge of foreign aid in a donor country faces the additional challenge in government budget negotiations that the direct beneficiaries of aid in the recipient countries are politically unimportant. Visible signs of success therefore become even more important to motivate the aid budget. This is reinforced by the fact that donor country constituencies typically have very limited knowledge of the size of the aid budget and how it is spent.[5] To have success stories and make them publicly known then becomes an important political objective to avoid aid fatigue. In this narrative, a solution to the problem thus entails a delegation of authority to independent aid agencies, shielded from political

[4] This is particularly true for aid-intensive countries with a shortage of educated and skilled workers, such as Mozambique, from which the following quote comes. "Donor-driven competition for skilled personnel is creating immense problems for government. The preoccupation of many donors with ensuring that their local administrations have a full complement of qualified staff and with securing, at all costs, the manpower required to implement their projects is depriving the government of the capacity to effectively manage its administration." (Fallon and da Silva, 1994, p. 98).

[5] For instance, public opinion polls in the US show that citizens typically overestimate the percentage of the federal budget devoted to foreign aid immensely (at a factor of 20-25 for the median response). As a consequence, most respondents saying the US government is spending too much on foreign aid, still suggest what they consider an appropriate aid budget that is vastly greater than the actual budget (PIPA, 2001).

pressure and thereby with a more long term perspective on development.

Prescriptions for how to allocate authority, and the meaning of accountability, thus differs across the two arguments. The first narrative accuses aid agencies of institutional inertia, and suggests that the solution is to give governments and tax payers better instruments to hold the agencies accountable. The second narrative puts the blame on political involvement, and instead argues that more independent agencies would prevent short sighted political considerations to creep in. Obviously, there is no allocation of control that can remedy both problems at the same time when neither government nor aid agency is believed to have only development in its objective function.

2.2 Impact Evaluation

Proponents of the first narrative point to the recent methodological "revolution", RCTs, as a potential remedy. These experiments have established a benchmark for how to quantify the causal effects of different types of interventions at the micro-level.[6] As such, RCTs offers a new and improved methodology for evaluating the effectiveness and efficiency of aid-financed projects. This in turn yields a better chance for governments and tax payers to hold aid agencies accountable for their actions. Proponents of this approach have gone as far as to argue that aid money should exclusively be channeled to projects that have been shown to have an effect through "hard" evidence, since results from alternative evaluation methods used in the social sciences are dismissed as lacking internal validity (Banerjee 2007). This criticism is also increasingly starting to have an impact on donors. In particular philanthropic organizations have become unwilling to finance projects that cannot be evaluated using RCTs, but also in the World Bank it is becoming the dominant tool for impact evaluation (Banerjee and Duflo, 2009).

[6] Interventions range everywhere from de-worming of school children, better access to textbooks or higher teacher-student ratios in schools, HIV-AIDS prevention strategies, information campaigns, to credit access within small firms (e.g. Miguel and Kremer, 2004, Glewwe et al. 2009, Duflo et al. (2006), Karlan & Zinman (2009) and Reinikka and Svensson, 2010).

The application of RCTs in evaluating development projects is a very important methodological advancement, with the potential to do good far beyond the academic world. In particular, by allegedly circumventing problems of selection on unobservables it helps with getting internally valid results of a causal impact. It is not above criticism, though. As pointed out in e. g. Deaton 2009, Ravallion 2009, Rodrik 2009 and Chen et al. 2009, identification still relies on certain assumptions that may be violated for instance in the case of spill-overs, unobserved heterogeneity in individual-specific program impact or unintended experimental impact on behavior of control groups or staff in the field.[7] Another concern is external validity. Replication studies in different environments have started to emerge, an important development to remedy concerns of lacking external validity, but so far results are mixed (e.g. Bobonis et al., 2006, and Duflo, Dupas and Kremer, 2008). So called general equilibrium effects from scaling up small scale programs have also been emphasized.[8] Acemoglu (2010) brings up the example of interventions that increase the return to schooling. When introduced locally, it may be plausible to think of the return to schooling as exogenous and independent of the intervention, but at a larger scale the skill premium is likely to go down if the project substantially increases the supply of skilled labor.

Of particular relevance for this paper is concerns about the impact on development policy if the possibility of RCTs becomes a requirement for financing. First of all, not all types of worthwhile interventions typically implemented by governments to fight poverty can be randomized. Macro-economic policies, infrastructure projects, public sector reform and institutional development, for instance, typically belong to this group of programs. Generally, the bigger and more complex the question, the harder it is to design a RCT to evaluate it. It follows that proponents of RCTs

[7] Note that conventional methods of empirical inference also suffer from many of these problems. The critique does thus not suggest that the methodology is necessarily inferior to existing methodologies, but it suggests that some caution may be warranted also when interpreting results derived under randomization, and that in practice traditional methods with more and better data may often do as well or better (Ravallion, 2012).

[8] There are a few examples of nation-wide experiments such as programs with mandated political representation by minorities in India (Chattopdhyay and Duflo, 2004, and Duflo and Topalova 2003).

typically also become proponents of a model of development based on small gradual policy reforms within the existing economic and political macro environment, underplaying the importance (or possibility) of more fundamental economic, political, social or attitudinal change (Banerjee and Duflo, 2011). This carries the risk that many potentially important interventions are never carried out because their effectiveness cannot be verified by "hard evidence". This problem is magnified by the fact that the bar for what is referred to as a clean identification is also continuously raised within the academic literature. In the end the methodology may end up determining what questions to ask, rather than letting an analysis of our knowledge gaps determine where to look for the answers.

Secondly, RCTs typically focuse on a subset of outcomes, those that are most easily observed and quantified. These outcomes may not be the most important ones, causing RCTs to "..produce clear answers to the wrong question." (Heckman 1992, 218). Components of learning and institutional development risk to become externalities if they cannot be quantified and packaged in the same way as say the number of children being vaccinated, the number of malaria bednets distributed or test scores in elementary schools. Excessive pressure to generate impressive immediate outcomes is a valid concern if learning and institutional development is important for long run sustainable development. This is not a critique of randomization as such, but it highlights that it is important to incorporate also measurement of factors relevant for long run development when performing impact evaluation. In particular, as illustrated in the model below, in a political environment in which immediate outcomes are already favoured, requirements on RCTs may help identifying the projects with the greatest short term potentials, but it also risks distorting *how* projects/programs are implemented.

3 The model

Think of a contractual agreement between a government (the principal) and an aid agency (the agent). Reforming the way aid is organized requires effort from the management of the agency so the government provides financial incentives to get aid more in line with its preferences. The contribution of aid to development depends on the quality of projects and programs but also the way in which aid is implemented. In particular I focus on the allocation of resources across operations and learning. Formally, I use the standard set-up of linear contracts, normally distributed noise and the constant absolute risk aversion (CARA) model of agent preferences (Holmström and Milgrom 1991). In the background is a function that maps the aid agencies actions into economic development, given by

$$Y\left(y_1\left(a_1, a_2\right), a_2\right), \tag{1}$$

where Y is the contribution of an aid budget of given size to economic development, y_1 is the immediately verifiable outcome of the projects or programs, $a_1 \in R^+$ catches the quality of the projects and programs in the portfolio, and $a_2 \in R$ represents the allocation of a fixed set of resources (humn and financial) between learning and operations. A higher level of a_2 implies more focus on operations, and thereby automatically less focus on learning. To keep the analysis simple, we assume the following specific functional form

$$Y = a_1 + a_2 + \varepsilon - \frac{\left(a_2 + 1\right)^2}{2}. \tag{2}$$

It is straightforward to show that maximizing the expected value of Y with respect to a_2 gives $a_2 = 0$. Thus, by construction, the optimal allocation of the resources is $a_2 = 0$, but note that this is just a normalization, it does not imply that the optimal amout of resources to one or the other is 0. A positive value of a_2 suggests that more resources are devoted to operations and less to

learning than what is optimal, while a negative value implies that too much resources are spent on learning and too little on operations. The parameter ε, drawn from a normal distribution with mean zero and variance σ^2, captures random outside factors that influence the success of projects. The expected value of y_1 is thus given by $a_1 + a_2$. Following the principal-agent literature, we interpret the variance of the underlying distribution as a measure of the ability to correctly evaluate the contribution to development from the design of the project/program rather than from luck, or other random case specific circumstances. In other words, a reduction in σ^2 means a better impact evaluation technology.

The utility of the risk neutral government is given by

$$\Delta - w, \tag{3}$$

where $\Delta \in \{Y, y_1\}$ and w represents the government resources channeled to the aid agency. The disutility from spending on aid comes from the opportunity cost of not spending those resources on something else of policy or political value to the government. The incentive contract offered to the aid agency is linear and given by[9]

$$w = \beta_0 + \beta_1 \Delta. \tag{4}$$

The utility of the (risk averse) head of the agency is given by

$$-e^{-\rho\left(w - \frac{1}{2}\left(c_1 a_1^2 + c_2 a_2^2\right)\right)}, \tag{5}$$

where $\rho > 0$ is the coefficient of absolute risk aversion.

The agency can improve project outcome through two means. The first way is to shift resources from existing less successful projects to new projects that have been shown successful elsewhere.

[9] The motivation for why the aid budget enters directly into the utility of the agency is that compensation, prestige and future career opportunities for management personnel typically depends positively on responsibility as measured by the size of the budget.

These changes are likely to meet resistance from aid contractors and personnel invested, financially or emotionally, in existing projects or programs, generating a tendency towards institutional inertia. Overcoming such resistance is costly, financially as well as in terms of effort. The marginal cost of project/program improvement is captured by the parameter c_1. The second way is by changing the allocation of resources between operations and learning to better reflect their relative contribution to development. This is also likely to be met by resistance, since it may include a shift in personnel, financial resources and focus. The marginal cost of such change is captured by the parameter c_2. Finally, the agencies reservation utility, in certain monetary equivalence since compensation from the contract is uncertain, is given by[10]

$$-e^{-\rho\hat{w}}. \tag{6}$$

The key informational assumption in the model is that the government can only observe the outcomes (Y, y_1), so the actual effort extended to identify and implement the best projects and the allocation of resources between operations and learning is private information. It follows that contracts can only be written based on outcomes.

3.1 The Patient Government

Think first of the case of a patient government that is not under immediate pressure to provide directly verifiable results to build political support. Such government will design an incentive contract that maximizes the total contribution of aid to development, i.e. $\Delta = Y$, subject to the opportunity cost of the financial resources and the incentive compatibility and participation constraints. The first thing to note is that the agency will set

$$a_2 = 0, \tag{7}$$

[10] This represents the minimum level of utility the agency management requires to stay in their jobs.

since that maximizes Y while at the same time minimizing the cost of reallocation of resources. The problem can thus be reduced to giving incentives to the agency to improve the quality of the projects, a_1.

Plugging equations (2) and (4) into equation (5) the incentive compatibility constraint can be written as

$$a_1 = \arg\max_{a_1} E\left(-e^{-\rho\left(\beta_0+\beta_1(a_1+\varepsilon)-\frac{1}{2}c_1a_1^2\right)}\right). \tag{8}$$

Solving the algebra yields the following unique solution

$$a_1^* = \frac{\beta_1}{c_1}. \tag{9}$$

Plugging in the values from equations (8) and (9) into the participation constraint yields

$$E\left[-e^{-\rho\left(\beta_0+\beta_1\varepsilon+\frac{1}{2}\frac{\beta_1^2}{c_1}\right)}\right] = -e^{-\rho\hat{w}}. \tag{10}$$

Using the fact that for any x and any normally distributed ε with mean zero and variance σ^2,

$$E\left[e^{x\varepsilon}\right] = e^{x^2\sigma^2/2}, \tag{11}$$

the participation constraint can be rewritten as

$$-e^{-\rho\left(\beta_0+\frac{1}{2}\frac{\beta_1^2}{c_1}-\rho\frac{\beta_1^2}{2}\sigma^2\right)} = -e^{-\rho\hat{w}}. \tag{12}$$

The value within the parenthesis on the left hand side is the certainty equivalent compensation of the agency, which includes a risk premium increasing in size with the coefficient of risk aversion, the variance of the output and the weight on output in the compensation contract. Solving for β_0 the participation constraint can be rewritten as

$$\beta_0 = \hat{w} - \frac{\beta_1^2}{2}\left(\frac{1}{c_1} - \rho\sigma^2\right) \tag{13}$$

Plugging in the derivations from the incentive compatibility and participation constraints, the government's program is now simplified to

$$\max_{\beta_1} \left(\frac{\beta_1}{c_1}\right)(1-\beta_1) - \hat{w} + \frac{\beta_1^2}{2}\left(\frac{1}{c_1} - \rho\sigma^2\right). \tag{14}$$

Working through the algebra yields

$$\beta_1 = \frac{1}{1+\rho\sigma^2 c_1}, \tag{15}$$

and

$$\beta_0 = \hat{w} - \frac{\left(1-\rho\sigma^2 c_1\right)}{2c_1\left(1+\rho\sigma^2 c_1\right)^2}. \tag{16}$$

Equations (7), (9), (15) and (16) summarize the incentive structure of the contract and the agent's response to those incentives. We can also calculate the expected contribution of aid to development as

$$E\left[Y\right] = \frac{1}{c_1\left(1+\rho\sigma^2 c_1\right)} - \frac{1}{2}. \tag{17}$$

The main results can be summarized as follows. A lower cost of overcoming resistance to change the project/program portfolio, a less risk averse head of the agency and a better technology for evaluation (a lower σ^2) will all contribute to a more high powered incentive contract, and a higher expected development impact of aid. In particular, when the government has patience and is not driven by short-sighted political considerations, then improvements in the ability to evaluate the effectiveness of aid projects (a lower σ^2) will help the government to hold the agency accountable and it will unambiguously have a positive effect on the contribution of aid to development.

3.2 The Impatient Government

Think now instead on the case in which the government is emphasizing immediately verifiable results, i.e. the case of $\Delta = y_1$.The incentive compatibility constraint can now be written as

$$\{a_1, a_2\} = \arg\max_{a_1,a_2} E\left(-e^{-\rho\left(\beta_0+\beta_1(a_1+a_2+\varepsilon)-\frac{1}{2}\left(c_1a_1^2+c_2a_2^2\right)\right)}\right). \tag{18}$$

Solving the above yields

$$a_1^* = \frac{\beta_1}{c_1}, \tag{19}$$

and

$$a_2^* = \frac{\beta_1}{c_2}. \tag{20}$$

Plugging in equations (19) and (20) into the participation constraint yields

$$E\left[-e^{-\rho\left(\beta_0+\beta_1\varepsilon+\frac{\beta_1^2}{2}\left(\frac{1}{c_1}+\frac{1}{c_2}\right)\right)}\right] = -e^{-\rho\hat{w}}, \tag{21}$$

which can be rewritten as

$$\beta_0 = \hat{w} - \frac{\beta_1^2}{2}\left(\frac{1}{c_1}+\frac{1}{c_2}-\rho\sigma^2\right). \tag{22}$$

The government's objective function is now given by

$$\max_{\beta_1}\left(\frac{\beta_1}{c_1}+\frac{\beta_1}{c_2}\right)(1-\beta_1) - \hat{w} + \frac{\beta_1^2}{2}\left(\frac{1}{c_1}+\frac{1}{c_2}-\rho\sigma^2\right). \tag{23}$$

Working through the algebra yields

$$\tilde{\beta}_1 = \frac{c_1+c_2}{c_1+c_2+\rho\sigma^2c_1c_2}, \tag{24}$$

and

$$\tilde{\beta}_0 = \hat{w} - \frac{\left(c_1+c_2-\rho\sigma^2c_1c_2\right)}{\left(c_1+c_2+\rho\sigma^2c_1c_2\right)^2}\frac{\left(c_1+c_2\right)^2}{2c_1c_2}. \tag{25}$$

Equations (19), (20), (24) and (25) summarize the incentive structure of the contract and the agent's response to those incentives. The main difference from the previous section is that the agency will now have incentives to reallocate resources away from learning, i.e. $a_2 > 0$. As long as the government cannot separate the effects of more effort on implementing better projects/programs, from that of spending more resources on operations and less on learning, then stronger incentives to bolster immediately verifiable output will distort resource allocation. The expected contribution of aid to development can now be written as

$$E[Y] = \frac{\tilde{\beta}_1}{c_1} - \frac{1}{2} - \frac{\tilde{\beta}_1^2}{2c_2^2}. \tag{26}$$

Compared to the case with the patient government, incentives are now stronger $\left(\tilde{\beta}_1 > \beta_1\right)$. It follows that more effort will be placed on overcoming resistance to change in the project/program portfolio. This will have a positive effect on the contribution to development, as captured by the first part on the right hand side in equation (26), since $\tilde{\beta}_1 > \beta_1$. This positive effect, though, is of second order to the negative effect on the allocation of resources across operations and learning, captured by the last term on the right hand side in equation (26). That the expected contribution to development is lower in this case is trivial, but what is more interesting is the impact of an improvement in the evaluation technology. Plugging in the solution from equation (24) into equation (26) yields the following expression for the expected contribution of aid to development.

$$E[Y] = \frac{c_1 + c_2}{c_1\left(c_1 + c_2 + \rho\sigma^2 c_1 c_2\right)} - \frac{1}{2} - \frac{\left(c_1 + c_2\right)^2}{2c_a^2\left(c_1 + c_2 + \rho\sigma^2 c_1 c_2\right)^2} \tag{27}$$

Taking the derivative of equation (27) with respect to σ^2, setting that derivative equal to or larger than 0 and solving for σ^2 yields a condition for when a better evaluation technology actually decreases the development impact of aid when the government emphasizes immediately verifiable

results.

$$\sigma^2 \leq \hat{\sigma}^2 \equiv \frac{(c_1 + c_2)\left(c_1 - c_2^2\right)}{c_2^3 c_1 p}.$$

Whenever the precision of the signal exceeds this threshold, a further improvement in the precision will decrease the contribution of aid to development. This happens because the negative effect on the allocation between operation and learning due to stronger incentives dominates the positive effect of better projects. Note that it is necessary that $c_2^2 < c_1$ in order for this condition to ever hold. Comparative statics reveal that the threshold level is generally increasing in c_1 and decreasing in c_2, i.e. the higher is the cost of changing projects/programs relative to changing between operation and learning, the larger is the span for which a more precise signal decreases the contribution of aid to development. To sum up the key implication of the model; not only does a solitary focus on the immediately verifiable output of aid decrease its development impact, but such focus can also cause better evaluation techniques to have the perverse effect of reducing aid effectiveness.

4 Implications

The mechanism outlined above suggests that in the presence of private information there may be a trade-off between motivating aid agencies to invest in the projects and programs with the highest immediate payoffs and making sure that local learning and institutional development is not neglected. Furthermore, no allocation of authority can solve this trade-off in an optimal way if aid agencies suffer from resistance to change their project/program portfolio whereas political concerns lead governments to focus on immediately verifiable results. Finally, in a second best world in which politicians focus on short term results, an advancement in project evaluation techniques will have the unintended consequence to decrease learning and institutional development. It is even possible that this effect dominates the positive effect on the ability to implement the best projects/programs,

suggesting that better tools can indeed make for worse outcomes, as suggested in the title of the paper. This comes about since the new methodology motivates a short sighted government to put stronger incentives on the aid agency to make sure projects are successful, which may lead to a reallocation of resources away from more long term goals. Note that the analysis doesn't ignore the potential benefits of better evaluations; any tool that helps to better identify what works and what doesn't obviously carries the potential to improve aid effectiveness. However, in a second best world, this tool may be difficult to direct precisely on the right target and it may be abused for purposes other than promoting long term development.

The result thus brings up the importance of taking political incentives in the donor countries into consideration before arguing that the "randomization revolution" should motivate a fundamental shift in foreign aid policies.[11] The first best solution is of course to have governments internalizing both learning and institutional development and immediately verifiable results. The focus on providing immediate success stories and, in particular, avoid negative media attention, is partly coming from ignorance among donor country citizens about the challenges of development. It is also fostered by competition between donors to show that their strategies indeed are generating results. In this context, international agreements between donors and recipients have the potential to work as commitment devices to avoid under-provision to the public goods of institutional development and learning. The actions decided upon in the Paris and Accra agendas with the purpose of channeling more resources through partner country financial systems, more emphasis on broad program support rather than narrow project support and reducing aid fragmentation are steps in the right direction. The extent of actual difference these agendas have made is less clear, though, as shown in for instance Wood et al. (2011). The agendas have no explicit punishment mechanisms for

[11] The concerns highlighted in this paper are not the only ones brought up in the debate on RCTs and their influence on development policy. See in particular Ravallion 2012 for a more complete list of potential concerns.

donors not abiding by the recommendations, and reports and evaluations from the OECD tend to be very unwilling to point the finger at anyone for not living up to expectations (e.g. OECD 2011). Their usefulness as commitment devices is thus very limited. As reported in for instance Birdsall and Kharas (2010), the continued use of PIUs also suggests that the temptation to prioritize project success is strong. The question is how much progress can be made only on the front of getting the incentives and patience on the political side right.

Maybe a better road to travel is to emphasize the importance of improved techniques for evaluation also of learning and institutional development, or more generally off "big questions". Some progress has been made to extend experimental approaches in this direction, for instance on political clientelism in Benin (Wantchekon 2003, Fujiwara and Wantchekon 2012). There are clearly limits on how far the RCT methodology can be taken though, so a more realistic approach to gain more knowledge of the big questions is to combine RCTs with structural econometrics and theory, using more and better data and game theory field experiments (e.g. Desai and Joshi, 2012). This would require that those in charge of aid policy admit the complexity of developmental work, and don't fall for the temptation to rely on just immediately verifiable results based on methods easy to explain to motivate what they are doing in the eyes of voters and tax payers. This may be critical to reduce the risk that the pressure to do the right thing makes us forget the importance of doing what we do the right way. That is, to make sure that domestic learning, institutional development and questions of long run sustainability are made an integral part of program and project implementation.

References

[1] Acemoglu, D. (2010), "Theory, General Equilibrium and Political Economy in development Economics", *Journal of Economic Perspectives 24*, 17-32.

[2] Acemoglu, D., S. Johnson and J. Robinson (2001), "The colonial origins of comparative development: an empirical investigation", *American Economic Review 5*, 1369-1401.

[3] Banerjee, A. (2007), "*Making aid work*", The MIT Press, Cambridge, Massachusetts.

[4] Banerjee, A and E. Duflo (2009), "The Experimental Approach to Development Economics", *Annual Review of Economics* 1, 151–78.

[5] Banerjee, A and E. Duflo (2011), "*Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*", PublicAffairs.

[6] Birdsall, N. (2004), "Seven Deadly Sins: Reflections on Donor Failings", CGD Working Paper 50.

[7] Birdsall, N. and H. Kharas (2010), "Quality of Official Development Assistance Assessment", CGD, Washington DC.

[8] Bjorkman, M. and J. Svensson (2009), "Power to the People: Evidence from a Randomized Field Experiment of Community-based Monitoring in Uganda", *Quarterly Journal of Economics* 124, 735-769.

[9] Bobonis, G., E. Miguel, and C. Puri-Sharma (2006), "Anemia and School Participation", *Journal of Human Resources* 41, 692–721.

[10] Chattopdhyay, R. and E. Duflo (2004), "Women as policy makers: Evidence from a Randomized Policy Experiment in India", *Econometrica*, 72 1409-1443.

[11] Chen, S., M. Ren, and M. Ravallion (2009), "Are There Lasting Impacts of Aid to Poor Areas?", *Journal of Public Economics* 93 (3–4): 512–28.

[12] Deaton, A. (2009), "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive keys to Economic Development", NBER Working Paper 14690.

[13] Desai, R. and S. Joshi (2012), "Collective Action and Community Development: Evidence from Women's Self-Help Groups in Rural India", Working Paper, Georgetown University.

[14] Duflo E, P. Dupas, M. Kremer, and S. Sinei (2006), "Education and HIV/AIDS prevention: evidence from a randomized evaluation in western Kenya", World Bank Policy Research Working Paper 402, Washington DC.

[15] Duflo, E., P. Dupas and M. Kremer (2008), "Peer Effects, Teacher Incentives, and the Impact of Tracking: evidence from a Randomized Evaluation in Kenya", NBER Working Paper 14475.

[16] Duflo, E. and P. Topalova (2003), "Unappreciated Service: Performance, Perceptions, and Women Leaders in India", Working paper, MIT.

[17] Easterly, W. (2006), "*The Elusive Quest for Growth* ", MIT Press, Cambridge, Massachusetts.

[18] Easterly, W. and T. Pfutze (2008), "Where does the Money go? Best and Worst Practices in Foreign Aid", *Journal of Economic Perspectives* 22, 29-52.

[19] Fallon, P., and L. Pereira da Silva (1994), "Recognizing Labor Market Constraints: Government-Donor Competition for Manpower in Mozambique". In D. Lindauer and B. Nunberg, eds., *Rehabilitating Government: Pay and Employment Reform in Africa*. World Bank, Washington DC.

[20] Fujiwara, T. and L. Wantchekon (2012), "Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence from Benin", Working Paper, Princeton University.

[21] Glewwe, P., M. Kremer, and S. Moulin (2009), "Many children left behind? Textbooks and test scores in Kenya", *American Economic Journal of Applied Economics* 1, 112–35.

[22] Heckman, J. (1992), "Randomization and Social Policy Evaluation." In Evaluating Welfare and Training Programs, edited by Charles F. Manski and Irwin Garfinkel, 201–30. Cambridge and London: Harvard University Press.

[23] Holmström, B. and P. Milgrom (1991), "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership and Job Design", *Journal of Law, Economics and Organization* 7, 24-52.

[24] Karlan D., and J. Zinman (2009), "Observing unobservables: identifying information asymmetries with a consumer credit field experiment", *Econometrica* 77, 1993–2008.

[25] Knack, S. and A. Rahman (2007), "Donor Fragmentation and Bureaucratic Quality in Aid Recipient Countries", *Journal of Development Economics* 83, 176-197.

[26] Miguel E., and M. Kremer (2004), "Worms: identifying impacts on education and health in the presence of treatment externalities", *Econometrica* 72, 159–217.

[27] OECD (2011), "Aid-Effectiveness 2005-2010: Progress in Implementing the Paris Declaration", OECD Publishing.

[28] PIPA (2001), "Americans on Foreign Aid and World Hunger: A Study of U.S. Public Attitudes", Program on International Policy Attitudes, University of Maryland.

[29] Ravallion, M. (2009), "Should the Randomistas Rule?", *Economists' Voice* 6.

[30] Ravallion, M. (2012), "Fighting Poverty one Experiment at a Time", book review in *Journal of Economic Literature* 50, 103-114.

[31] Reinikka, R., and J. Svensson (2010), "Working for God? Evidence from a change in financing of not-for-profit health care providers in Uganda", *Journal of the European Economic Association* 8(6),1159–1178.

[32] Rodrik, D. (2009), "The New Development Economics: We Shall Experiment, but How Shall We Learn?" In What Works in Development: Thinking Big and Thinking Small, edited by Jessica Cohen and William Easterly, 24–47. Washington, D.C.: Brookings Institution Press.

[33] Rodrik, D., A. Subramanian and F. Trebbi (2004), "Institutions rule: The Primacy of Institutions over Geography and Integration in Economic Development", *Journal of Economic Growth* 9, 131-165.

[34] Sachs, J. (2005), "*The End of Poverty: Economic Possibilities for Our Time*", Penguin Press.

[35] Seabright, P. (2002), "Conflicts of Objectives and Task Allocation in Aid Agencies: General Issues and Applications to the European Union", in B. Martens, ed., *The Institutional Economics of Foreign Aid*, Cambridge University Press, Massachusetts.

[36] Wantchekon, L. (2003), "Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin", *World Politics* 55, 399-422.

[37] Wood, B., J. Betts, F. Etta, J. Gayfer, D. Kabell, N. Ngwira, F. Sagasti, and M. Samaranayake (2011), "The Evaluation of the Paris Declaration, Final Report", Copenhagen.