

## **Assessing the Efficacy and Institutional Impacts of Cash on Delivery aid: The Necessity but Insufficiency of a Demand-Side Emphasis**

Michael Woolcock, University of Manchester<sup>1</sup>

February 15, 2008

(First rough draft; please do not cite)

### Abstract

Recent initiatives such as ‘Cash on Delivery aid’, in which demonstrated progress towards the attainment of development goals (e.g., primary school enrolment)—as opposed to satisfying, *ex ante*, various bureaucratic or political ‘conditions’—becomes the basis for receiving foreign aid, are to be welcomed and encouraged. If the overriding objective, however, is to enhance the efficacy (and, by extension, the volume and impact) of aid, then a shift from a ‘supply side’ to ‘demand side’ focus is necessary but insufficient. A truly more rigorous understanding of whether and how development assistance ‘works’, which can in turn be a basis for more effective learning and decision-making by both donor and recipient governments, requires an appropriate match between the known shape of benefits flows over time—at the level of the sector (e.g., education) and projects/policies designed to attain it (e.g., conditional cash transfers)—and a corresponding array of research tools capable of discerning such a flow. At present, however, the development community knows very little, other than by assumption, about how the expected shape of the benefits flow from *any* given sector or project type (roads, bridges, irrigation, health, ‘empowerment’) changes not only over time, but also as a function of scale, context, and implementation effectiveness, and as such is prone to routinely making Type I and Type II errors; these concerns are only compounded when development researchers become wedded (as too many have) to a very narrow methodological basis (i.e., randomized trials) for determining project effectiveness. Cash on Delivery aid partially addresses (though does not solve) some of these concerns, and as such should be part of, not a substitute for, a broader strategy for enhancing the volume, quality and allocation of aid.

---

<sup>1</sup> Research Director, Brooks World Poverty Institute (on external service leave from the Development Research Group at the World Bank). The views expressed in this paper are those of the author alone, and should not be attributed to the University of Manchester or the World Bank. Preliminary discussions on these issues with Michael Clemens, Elizabeth Levy Paluck, Lant Pritchett, Vijayendra Rao, Sandra Sequeira and Martin Ravallion have been most helpful (though none should be implicated where there are errors of fact or interpretation). Address for correspondence: Brooks World Poverty Institute, University of Manchester, Humanities Bridgeford Street Building, Oxford Road, Manchester M13 9PL, United Kingdom. Email: [michael.woolcock@manchester.ac.uk](mailto:michael.woolcock@manchester.ac.uk)

## I. Introduction

Contemporary discussions regarding the efficacy of foreign aid are dominated by high-profile (and often quite acrimonious) debates between so-called “optimists” (Sachs 2005) and “pessimists” (Dichter 2003, Easterly 2006), with others (Collier 2007) seeking to occupy the “sensible center”<sup>2</sup>. The basis for the key distinctions between these camps has been somewhat crudely identified as the extent to which large-scale, top-down, external, expert-driven, technocratic inputs (overseen by “planners”) can successfully break the many and various “poverty traps” that keep poor people (and/or poor countries) poor, or whether the preferred starting point should instead be more modest, bottom-up, entrepreneurial, experimental initiatives built on local knowledge in response to context-specific problems (piloted by “searchers”), with larger projects only undertaken once rigorous evidence has established an initial positive net impact.

In the context of widespread popular ignorance about the scale of foreign aid<sup>3</sup> and enduring scepticism regarding the extent to which it yields positive and “sustainable” impacts, these debates serve a useful purpose by raising awareness, clarifying the nature and basis of diverse perspectives, and encouraging more hard-nosed efforts to understand “what works and why” (World Bank 1998, Riddell 2007). Even so, however, the current terms of debate—at least as articulated by each camp’s most strident protagonists and ardent followers—inadequately capture (and thereby “crowd out”) important alternatives being explored by innovative development professionals and concerned citizens alike. Moreover, it is not clear that “development” is always and everywhere a “problem” for which a corresponding policy or project “solution” (of whatever kind) can be identified, implemented, and empirically assessed, or that the structure and imperatives of the international aid architecture—originally established to attend to less ambiguous development problems (such as ‘reconstruction’)—is adequately set up to fulfil its core mission. Nonetheless, for better or worse, these are the architectures and institutions we have, and we must work with them if we are to improve the quality (and in time the quantity) of the aid they oversee.

One such innovative being pursued by development professionals—in particular, the Center for Global Development—is an initiative known as Cash on Delivery aid (hereafter COD). With COD, the emphasis is not, as with traditional aid efforts, on ensuring *ex ante* commitment by the recipient government to a list of ‘conditions’ (e.g., “good governance”, allocation of resources to predetermined sectors or contractors, etc) as a requirement for receiving aid, but an empirical demonstration (determined via objective, verifiable indicators) of incremental improvement towards agreed-upon goals, such as increased primary school enrolment, lowering the incidence of malaria, etc (see Barder and Birdsall 2006). Through a combination of establishing clear performance metrics and incentives, deploying powerful peer and demonstration effects, providing more and better public information, and using concrete development achievements rather than vague adherence to ‘conditions’ as the dominant accountability mechanism, the hope is that (a) local service providers will be better able to identify and implement solutions to local problems, (b) all actors in the delivery chain (from donors to national governments to front-line

---

<sup>2</sup> See Clemens (2007) for a useful critique of these various positions.

<sup>3</sup> Pollsters repeatedly find that citizens of OECD countries vastly overestimate the amount of money (as a percentage of the federal budget) given by their government for the purpose of foreign aid.

service providers) will face stronger pressures to serve the public (rather than their own private) interest, and thus (c) that ultimately the primary intended beneficiaries—i.e., the poor—will enjoy a greater share of the actual and potential fruits of development. This is an important initiative, and it should be trialled and assessed on its merits to see whether, compared to plausible alternatives (i.e., the counterfactual by way of either ‘nothing’ or aid volume and delivery modalities that would prevail were COD not to be in place), it does indeed achieve a net positive impact.

Will COD achieve these goals? I argue that this question must be answered at two levels, the first in terms of whether COD is likely to be an improvement over prevailing practices (which I argue it will), and the second in terms of whether COD alone can resolve the larger and more important identification problem, namely of accurately discerning—for the purposes of encouraging learning, efficiency and effectiveness—whether development resources allocated to sectors, and projects striving to achieve the stated goals of a given sector, are in fact driving outcomes in the desired direction (which I argue it cannot). As such, this paper seeks to do two things. First, to identify and assess the various institutional constraints against which COD advocates might have to contend in taking their ideas forward, from the drawing board to government agencies to implementing organizations to the poor themselves. Second, it seeks to clarify the nature of the broader conceptual (but no less practical) issues that must be addressed in order to more accurately assess the efficacy of aid. If this is in fact the key overriding objective of the COD initiative, the paper argues that merely shifting the onus of responsibility for aid effectiveness from the ‘supply side’ to the ‘demand side’ overlooks a central but largely unknown aspect of most development sectors, policies and projects, namely their “functional form”, or how their impacts change over time, context, scale and implementation effectiveness. Without such knowledge, even the cleanest implementation strategy and the most rigorous assessment protocol will be unable to make definitive statements about whether and how impact is being achieved, and thus whether calls to expand and/or replicate a given project (or assistance modality more generally) are justified. In the second part of the paper, I thus outline the nature and seriousness of the functional form problem, and how COD advocates might address it as part of their broader campaign to improve the quality and quantity of aid.

## **II. Assessing the Institutional Impact of Cash on Delivery aid**

Low-income countries vary considerably with respect to the extent to which ‘foreign aid’ is a significant component of their national economic life; as such, an obvious starting point when considering reform possibilities is to recognize that individual poor countries—e.g., India (low) versus Zaire (high)—face very different incentives to modify their behavior in response to any effort by donors to ‘rethink’ assistance strategies and/or modalities. Moreover, the fungibility of aid within the larger ambit of a country’s budget means that ‘reforms’ in any particular sector or ministry may or may not yield higher aggregate efficiency. Nevertheless, to the extent that the median aid recipient country has an active interest in securing an ongoing stream of external resources and that it is serious about allocating those resources for development purposes, it is worth exploring whether and how orthodox donor assistance procedures can be reworked to achieve greater effectiveness, and what institutional challenges—even when the prevailing imperatives seemingly favor change—might stand in the way of reform.

Cash on Delivery aid represents one such approach to reworking traditional aid delivery mechanisms, in which the emphasis over many decades has been on adherence to upholding ex ante conditionality agreements. COD endeavors to reverse this emphasis, instead asking countries to demonstrate verifiable progress towards the attainment of development goals that the recipient countries themselves have set as the basis for receiving ongoing financial assistance.

From an institutional perspective, the key issue (I think) is whether COD's reworking of aid incentives will significantly help move the core service delivery agencies in recipient countries from a 'pre-modern' to (something more closely resembling) a 'modern' institutional form—that is, from an agency likely characterized by clientelism, patrimony and weak accountability to one functioning along more rules-oriented and clear performance-based (i.e., 'professional') lines. Although COD advocates are, in principle, less concerned with the details of whether and/or how such transitions occur, so long as demonstrable outcome improvements are achieved, it is surely most likely to be the case that *sustainable* improvements in performance outcomes will occur in the first instance by refining the exiting institutional system but only in the 'second instance' (and beyond) by qualitative shifts in the organizational infrastructure responsible for such delivery. Moreover, to the extent there are indeed "lessons" for institutional design of broader significance that emerge as a result of implementing COD contracts, articulating those lessons will require some knowledge of the processes that yielded them—that is, procuring knowledge of what went on inside the 'black box' that brought about the change will be as important (to recipient as well as donor countries) as the brute fact of whether or not 'goals' were attained.

More formally assessing the nature and extent of any institutional changes associated with introducing COD will require a careful research design. There are three basic options: (a) a comparison of two similar countries seeking a similar development objective (in, say, education), one with and one without COD; (b) a comparison over time of a particular country's development performance in that sector, juxtaposing the experience of an earlier period (presumably over many years) with 'traditional' approaches with a subsequent period deploying COD; and (c) an intra-country (or even intra-state) comparison in which different (but otherwise similar, matched on observable and 'unobservable' characteristics, but spatially separate to avoid 'spillover' effects) regions agree to simultaneously implement 'orthodoxy' and COD in the same sector. There are of course many logistical, political and ethical reasons why implementing all or any of these institutional assessment protocols may not be feasible, but (other things equal) the third option above is likely to be one that yields the most fruitful insights, since it 'controls' most readily for national-level institutional factors that may be driving observed outcomes. As a matter of principle, I believe it is also the case that the lessons for developing countries from their own internal experiences with service delivery-type goals (e.g., health, education, water) trump those that may emerge from other countries.

To this end, and to most deftly understand the various institutional impacts of COD, I would propose working with an aid recipient country that is willing and able to implement COD in a select number of sub-national areas. It is highly unlikely that a COD-type initiative could be "randomly" assigned across a country, so the next best alternative is to use a double difference matching strategy, in which baseline and outcome data are collected on both COD states/districts and non-COD states/districts that are as otherwise similar (demographically, geographically, politically) as possible. Crucially, however—as will be discussed in more detail in the next

section—it will be vital to collect and track, over time, data on institutional *processes* that accompany the roll-out of COD. In addition to collecting the quantitative performance data, this will necessarily entail carrying out insightful qualitative research on whether and how COD changes such things as the nature of decision making at different points in the service delivery chain<sup>4</sup>, how realigned incentives (COD’s central innovation) at the ‘center’ alter the behavior of front-line service providers, and—perhaps most important of all—how COD alters (if at all) the various “relationships of accountability” (World Bank 2003) between line ministry officials, service providers, and clients (i.e., poor people) with respect to their commitment to achieving agreed-upon development goals.<sup>5</sup> (Indeed, learning more about how the goals themselves came to be “agreed upon” would also be really interesting and useful.) Finally, such qualitative process research would be vital for exploring the dynamics of changes in allied organizations—such as teachers unions, medical associations, and rival political parties—whose support for (and/or resistance to) service delivery innovations is clearly central to realizing COD’s objectives.

How might this type of research actually be done? First, it is vital that research (qualitative and quantitative) be explicitly incorporated into the contract that underpins the COD agreement; donors and recipients must agree that, if genuine learning for all parties is to occur and if key decisions going forward are to be based on the most rigorous and comprehensive evidence base possible, having a strong research component is in everyone’s interest. Second, there needs to be explicit commitment to the importance of understanding process issues, to recognizing that collecting quantitative data at the ‘beginning’ and ‘end’ of the exercise is very necessary but very insufficient for credibly discerning both the kinds of institutional changes that are unfolding *and* (most importantly; see next section) whether indeed the entire COD exercise is having the kind of positive impact it hopes it will. Third, the research teams to do this work will need to be not only carefully selected but also actively trained (probably over several weeks) in the methods and analytical procedures required. The extent of these efforts will depend on the country context, which may or may not have a plentiful supply of solid social science researchers, but even if it does the quality of the end product will be a direct function of the efforts expended up front on training and team building. Fourth, researchers—especially the qualitative researchers—will need to have the freedom to regularly interview (and where appropriate, observe) key actors in the process and to uphold confidentiality agreements with informants. Having ready access to project monitoring data will also be important.

With adequate political support from the outset and a shared commitment to implementing as rigorous a research design as is feasible, it should be possible, in principle, to generate more and better knowledge to help everyone understand whether and how COD is working, and how it is influencing the key institutions responsible for service delivery in the country in question. Given the lives and resources that are at stake in the foreign aid business, however, accurately determining the impact of project, policies and delivery mechanisms takes on additional significance. Unfortunately, however, the development fraternity largely lacks important

---

<sup>4</sup> One could see this as a more refined qualitative attempt to unpack the flow of resource championed by advocates of ‘expenditure tracking surveys’.

<sup>5</sup> The breakdown in relationships of accountability was identified by the World Bank (2003) as one of the key factors underpinning the routine failure of service delivery sectors such as education and health, where the principal-agent problem—that is, where frontline workers (teachers, nurses) typically operate far removed from oversight by their immediate superiors—is so acute.

knowledge of how the trajectory of impacts associated with key sectors (public health) and projects within them (AIDS awareness) unfolds over time, and how these trajectories of impact change as a function of context, scale and implementation effectiveness. In the section that follows, I outline the seriousness of this problem and the challenges that COD must confront as it seeks to substantiate its efficacy vis-à-vis plausible alternatives.

### III. Why Knowledge of a Project’s “Function Form” Matters for Assessing Impact

In discussions of how to enhance the efficacy of aid, most of the emphasis at present is on how to upgrade the ‘rigor’ of the tools and techniques used to assess impact—i.e., on *methodological* aspects of the *evaluation* protocol itself. Even in moments of critical self-reflection, advocates of ‘gold standard’ techniques (i.e., randomization) (Banerjee 2007) couch their concerns primarily in terms of the ethics, logistics or political feasibility of assigning actual or potential participants to ‘treatment’ and ‘control’ groups. At the end of the day, however, there remains the strong belief that gold standard protocols are the most (and for true believers, *only*) defensible basis on which to make hard decisions about whether a given project or policy is ‘working’ and thus whether (or not) it should be continued (terminated), scaled up (down) or replicated elsewhere.

The task of this section is to highlight the importance of striving for a better understanding of the *theory* of the *project’s* known or likely impact over time, context, scale and implementation effectiveness; indeed, I argue that without such knowledge, it is virtually impossible to say anything meaningful about project impacts, irrespective of whether they are inspired by ‘demand’ or ‘supply’ side imperatives, and even if they have been subject to the most deftly implemented randomized trial. Specifically, I argue that a truly rigorous evaluation is one that deploys the best available assessment tools at intervals that correspond to the shape of a project’s known (experientially, empirically or inferred on the basis of sound theory) impact over time. For the purposes of simplicity of exposition I shall call the shape of this net benefits profile over time the project’s or policy’s “functional form”—that is, the underlying “technology” that forms the core essence of the project and that is indeed truly independent of scale, context and implementation effectiveness.<sup>6</sup>

As things currently stand, the functional form for virtually all development projects—from rural roads and urban sanitation to guaranteed work programs and microfinance initiatives—is effectively assumed to be (net of everything else) monotonically increasing and linear (i.e.,  $y = mx + c$ , as in the dashed line in Figure 1 below<sup>7</sup>), with the only serious empirical issue left to determine being the difference between the benefits obtained by the project participants and the counterfactual (‘control’) group, net of both groups’ “initial conditions” (this is the classic

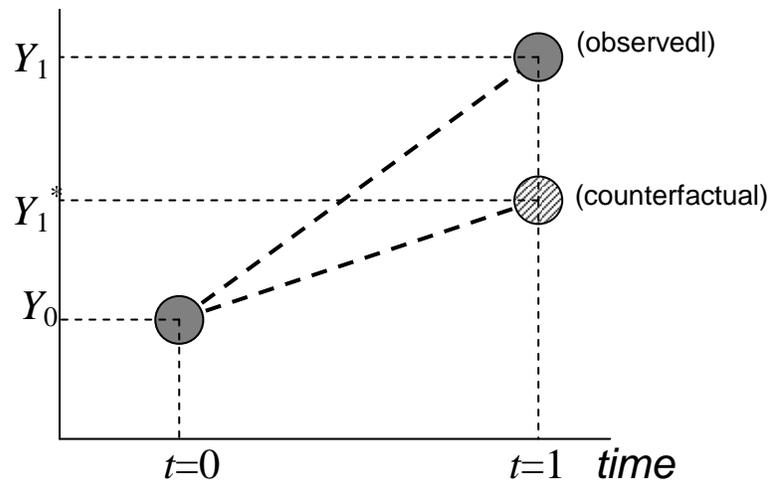
---

<sup>6</sup> In certain respects, this paper thus seeks to extend to the sectoral and project level the logic of the arguments developed by Pritchett (2000) for understanding the trajectories of economic growth, the “shape” of which manifestly varies enormously (though this is under-appreciated) over time and across countries. The key difference is that economic growth is clearly a dependent variable (i.e., an ‘end’) whereas most projects and policies are presumed to be independent variables (‘means’) exhibiting some degree of stability across time and space. I argue below that such assumptions, particularly in sectors or projects that necessarily entail high degrees of front-line discretion and transaction intensity (Pritchett and Woolcock 2004) have little empirical or theoretical basis.

<sup>7</sup> There are some important exceptions (e.g., Moffit YEAR), but it remains the case that the dominant assumption in the vast majority of discussions about aid (policy, project) effectiveness assumes a linear functional form.

‘double difference’ protocol). Put another way, when  $t_1$  is actually calculated is largely irrelevant: with a functional form that is assumed to be monotonically increasing and linear, the net benefits to project/policy beneficiaries will be deemed to be “significant” at the magical statistical moment when the difference between  $Y_1$  and  $Y_1^*$  passes the 0.05 level<sup>8</sup>; moreover, such an assumption allows project managers and policymakers to extrapolate the flow of benefits (far) into the future.

**Figure 1: The canonical basis for assessing project/policy impact**



But it is only the most ad hoc theorizing or wishful thinking that could possibly substantiate such an assumption and the evaluation protocols to which it gives rise (that, or the overriding imperatives of domestic political cycles and/or the nature of career paths at development organizations). I argue that for many (if not most) of the projects in its repertoire, the development fraternity actually has a very weak substantive understanding, *ex ante* or even *ex post*, of what the functional form of those projects is or might be: we know we need ‘baseline’ data (at time  $t_0$ ) and follow-up data (at time  $t_1$ ), but the proverbial ‘black box’ connecting them remains wholly a mystery, to the development industry’s peril. This is especially true of ‘social’ projects, such as those trying to increase the participation of marginalized groups, enhance women’s empowerment or access by the poor to formal legal systems. Even a cursory reading of social theory, for example, would suggest that in fact the most likely shape of such projects’ functional form is a J-curve (i.e., things get worse before they, hopefully, maybe, get better) or a step function (i.e., long periods of stasis followed by a sudden rupture brought on by, say, the reaching of a “tipping point” in the adoption of a new fertilizer technology, in which prevailing norms and/or uptake by an influential local leader suddenly but universally leads others to do likewise).<sup>9</sup> Moreover, one can imagine that, in reality, all manner of functional forms actually

<sup>8</sup> I leave aside here—though the thrust of my remarks on this point are surely consistent with—McCloskey’s lifelong quest to implore economists (and by extension other social scientists) to appreciate the difference between statistical and economic significance (see, only most recently, Ziliak and McCloskey 2008).

<sup>9</sup> Related arguments are made in the field of evolutionary biology, for example, where Gould (YEAR) famously (if controversially) argues forcefully for the notion of ‘punctuated equilibrium’—species stay the same for long periods of time, then change rapidly in response to an external shock or a mutant gene (e.g., changes in pigmentation) that

characterize the different types of interventions that comprise ‘development’ today: in addition to the above examples, anti-corruption efforts (e.g. public expenditure tracking surveys) may have high initial impact that fades over time (as the bad guys figure out alternative ways to cream off funds); etc. Appendix A offers some prognostications on the possible shape of the impact curve associated with various different types of development interventions.

Development theorists, evaluators, project managers and senior administrators all need to be explicit in articulating and substantive in defending the functional form that they believe (know?) characterizes their interventions. The absence of such knowledge can lead to potentially major errors (Type I and II) in assessing the efficacy of projects: if a women’s empowerment project has a J-curve functional form (men initially resist ceding resources and status, perhaps violently at first, only to come around when their attitudes and/or interests change), and a ‘gold standard’ randomized evaluation happens to be conducted when the program is at the bottom of that curve, it would be unceremoniously (but very inaccurately) deemed a failure. Similarly, projects to enhance human rights may take decades (as did, for example, efforts to end slavery) to be achieved, and vastly outlive the career of a single project manager. In a world that rewards rapid and easily measured development gains, managers of such projects are likely to be promoted long before their equally diligent and talented colleagues slogging away faithfully on projects that are ‘intrinsically’ important and ‘demand driven’ but years away from demonstrable ‘instrumental’ results. When coupled with international mandates to achieve “targets” such as the MDGs, the net effect is that the entire development industry ends up ‘reverse engineering’ itself, strongly preferring “high initial impact” projects over projects that actually respond to problems that poor countries themselves deem a priority but which are hard to measure and/or slow to demonstrate impact.

Either way, more rigorous knowledge of any given development project’s functional form is needed as a vital complement to efforts to enhance the quality of the tools deployed to assess them. Indeed, without such knowledge, we will be in no stronger a position to make hard decisions about whether and how projects are ‘working’.<sup>10</sup>

#### **IV. Implications and Conclusions**

Proposals to shift the dominant emphasis in foreign aid from ex ante ‘conditionality’ adherence to ex post performance attainment represents an important advance in the political economy of aid and a welcome (if belated) recognition that “getting the incentives right” must be a central principle of efforts to improve the quality and quantity of aid. If the overriding objective of such proposals, however, is to improve the basis of learning and decision-making for both donors and recipients of aid, then it is vital to obtain knowledge about the expected nature of the trajectory of the impact characterizing a particular sector (e.g., education) over time and individual projects designed to meet that sector’s objectives. If Clemens (2004), for example, is correct that the

---

proves qualitatively superior in addressing a particular problem (avoiding predators). See also Thomas Kuhn’s (1968) (even more famous and controversial) arguments that “paradigm shifts” in science (and elsewhere) follow a similar step-function trajectory, with orthodoxy entrenching itself and then prevailing well beyond its time as an old guard initially successfully defends it but then dies off, with a new paradigm then rapidly taking its place.

<sup>10</sup> Space constraints preclude a more detailed discussion of these issues; see Paluck and Woolcock (2008) for such a discussion and some constructive suggestions for how to respond to them.

shape of school enrolments over time is consistently an ‘S’ curve, then the first task for education policymakers in a given country is to identify where their country is (and regions within it are) placed on that curve, the better to understand where the greatest marginal gains are to be had (and/or to help mute unduly high expectations that low initial enrolments can be easily expanded); the second (and potentially more vexing) task is to identify specific implementation mechanisms—such as, say, conditional cash transfer programs—that are most likely to help with such efforts, but to also understand that these mechanisms themselves contain a “functional form”—that is, their impact over time, in different places, at different scales, and in accordance with how well they are implemented, are likely to vary enormously. Knowledge of exactly how, where and when this variance manifests itself is crucial for making accurate empirical evaluations of project/policy effectiveness. As such, assessing the recipients of aid assistance according to whether they have achieved their stated “goals”—and using such assessments as the basis for determining further streams of funding—can only be meaningfully undertaken, I contend, when knowledge of the functional form of a given sector and particular projects/policies is present. Absent such knowledge, the capacity for fundamentally inaccurate conclusions to be drawn (or, more formally, for Type I and Type II errors to be made) is rife.

There is a final danger of taking the seemingly pragmatic approach to aid delivery of rewarding demonstrated success, namely that it will either discourage the very experimentation (which, by definition, implies regular failure) needed to identify solutions to development problems. The response to this concern from COD advocates may be that, with appropriately designated time horizons, the process of experimentation can play itself out long enough for effective solutions to emerge, but even if such space could be created the basic imperatives of any objectives-based approach (which includes the Millennium Development Goals) almost inevitably must mean that any such experimentation is the exception rather than the rule. Carried to its extreme, the danger is that the entire development enterprise ends up being “reverse engineered”, with donor and recipient governments alike strongly favoring “proven technologies”, irrespective of the idiosyncrasies of the particular development problems that prevail in a particular country—that is, that “solutions” will end of determining the problems, rather than (as it should be) the other around.

As such, COD should be part of, rather than a substitute for, an approach to reformulating development strategies, a reformulation that stresses not only true “country ownership” and a correct alignment of incentives for all actors involved, but also a more coherent sequencing of deliberations from identification of problems to exploration of possible solutions to knowledge of how those solutions are likely to change as a function of time, scale, context and implementation effectiveness. Acquiring such knowledge will not be a product of simply deploying what some deem to be “gold standard” evaluation protocols (e.g., randomized trials), but rather deep engagement with project and policy processes, calling upon the full arsenal of research tools (qualitative and quantitative) available to social scientists.

## References

- Barder, Owen and Nancy Birdsall (2006) "Payments for progress: a hands-off approach to foreign aid" Working Paper No. 102, Center for Global Development, Washington, DC
- Banerjee, Arbjhit (2007) *Making Aid Work* Cambridge, MA: MIT Press
- Clemens, Michael (2004) "The long walk to school: international education goals in historical perspective" Working Paper No. 37, Center for Global Development
- Clemens, Michael (2007) *Foreign Affairs*
- Collier, Paul (2007) *The Bottom Billion* New York: Oxford University Press
- Dichter, Thomas (2003) *Despite Good Intentions: Why Development Assistance to the Third World Has Failed* Amherst, MA: University of Massachusetts Press
- Easterly, William (2006) *The White Man's Burden*
- Easterly, William (2008) "How the millennium development goals are unfair to Africa" Working Paper No. 14, Global Economy and Development Program, Brookings Institution
- Gould, Stephen Jay (2007) *Punctuated Equilibrium* Cambridge, MA: Harvard University Press
- Kuhn, Thomas (1968) *The Structure of Scientific Revolutions* Chicago: University of Chicago Press
- Moffit, Robert (YEAR)
- Paluck, Elizabeth Levy and Michael Woolcock (2008) "On identifying and responding to the 'functional form problem' in the assessment of development projects" Mimeo
- Pritchett, Lant (2000) "Understanding patterns of economic growth: searching for hills among plateaus, mountains and hills" *World Bank Economic Review* 14(2): 221-50
- Pritchett, Lant and Michael Woolcock (2004) "Solutions when the solution is the problem: arraying the disarray in development" *World Development* 32(2):
- Riddell, Roger (2007) *Does Foreign Aid Really Work?* New York: Oxford University Press
- Sachs, Jeffrey (2005) *The End of Poverty* New York: Penguin
- World Bank (1998) *Assessing Aid: What Works, and Why* New York: Oxford University Press
- World Bank (2003) *World Development 2004: Making Services Work for Poor People* New York: Oxford University Press
- Ziliak, Stephen and Deirdre McCloskey (2008) *The Cult of Statistical Significance: How the Standard Error Cost us Jobs, Justice and Lives* Ann Arbor: University of Michigan Press

**Appendix A: Some possible shapes of impact curves for various development projects**

