WP 2012-03
January 2012

Working Paper

Charles H. Dyson School of Applied Economics and Management
Cornell University, Ithaca, New York 14853-7801 USA

RESULTS BASED DEVELOPMENT ASSISTANCE: Perspectives from the South Asia Region of the World Bank

Thomas O'Brien, Ariel Fiszbein, Alan Gelb, Ravi Kanbur, and John Newman

Results Based Development Assistance: Perspectives from the South Asia Region of the World Bank

Thomas O'Brien, Ariel Fiszbein, Alan Gelb, Ravi Kanbur and John Newman

December, 2011

Contents

1. Results Based Development Assistance and South Asia

Thomas O'Brien, Manager for Development Effectiveness
South Asia Region, The World Bank

1. Introduction

The World Bank's mission to reduce poverty across the globe is a powerful motivator which inspires and guides our work. Indeed we came into development to make a difference over the long term and to have a measurable impact on development *outcomes*. And yet a fair amount of the day-to-day work that we undertake, and the operational and financial frameworks within which we operate, is centered on *inputs*. Operational staff in development agencies, together with our partner countries are often frustrated about this focus on monitoring and managing inputs, especially the paperwork that surrounds these tasks. But what if we could turn things on their head, and instead, relate development finance more closely to outcomes achieved rather than to inputs used?

This perspective on development finance, sometimes known as Results Based Financing (RBF), is now being discussed actively in development agencies and in the development community more broadly. The South Asia Region of the World Bank (SAR) has shown great interest in the concept, and has pioneered attempts at implementation. With this as background, the South Asia Region Development Effectiveness Unit (SARDE) organized a major workshop to gather different perspectives on RBF, to inform and explore this fascinating subject.

This compendium brings together four key presentations at the workshop, by Ariel Fiszbein, Alan Gelb, Ravi Kanbur and John Newman which serve to frame the broad issues, and which give examples of success and also delineate the challenges.[1] As a prelude to the four presentations which follow, this brief overview introduces the key issues and challenges in results based development assistance as they emerged in the workshop discussions.

2. Key Issues

The Results Chain

Most of our speakers grounded their analysis in a review of the results chain. Ravi Kanbur emphasized that if a results-based focus is to mean anything, then all projects must specify outcome, how it will be measured, and how the outcome will be attributed to the project inputs. John Newman drew out this point further with two powerful examples – one from a

[1] The conference sessions were chaired by Xian Zhu, SAR's Director of Strategy, and Kalpana Kochhar, SAR's Chief Economist. We were joined by Paul Bermingham, Director of the World Bank's Operational Services Department, to discuss recent developments in advancing the results based approach in the World Bank's lending. Preeti Arora, Senior Economist, SARDE, presented an overview of inputs from SAR field offices at the workshop. Fayez Omar, at that time the Director of Operational Services, and Thomas O'Brien, Manager for Development Effectiveness, helped facilitate the discussions.

state-wide public sector performance initiative in Minas Gerais (Brazil), and the other in tackling malnutrition, a truly multisectoral issue.

Alan Gelb focused attention on the fact that an important point of any results chain to work in practice is that there must be a decent "feedback loop", so that as performance is monitored, it can have some impact on actions. He argued that in many of the more traditional Bank projects, the feedback loop is too long and protracted to really be of much use (and cited the archetypal 54 year old Bank task team leader who likely will be retired before any evaluation of project outcomes is ready!).

Both Fayez Omar and Thomas O'Brien acknowledged these issues and explained how they are, in fact, ever-present in the work of any large scale development agency including the World Bank. O'Brien argued strongly that the results-based approach, however, is a powerful lever to break through some of the weaknesses of the past, since it much more centrally and pivotally requires proper results chains and results measurement to work in practice.

The Applicability of the Results-Based Approach

None of the speakers claimed that a results-based financing is a magic bullet, and all cautioned that it has to be used in circumstances where it fits. Ariel Fiszbein brought forward an interesting assessment of how, certainly in the social sectors, there is a "middle-ground" (in topics such as improving public service delivery) where neither the Bank's traditional investment operation nor a policy-based operation fits the bill. Here there is a great scope for RBF. Indeed Thomas O'Brien pointed out that SAR already has two highly successful results-based operations in the education sector—in Punjab and Sindh in Pakistan—which fall in this category.

Ravi Kanbur discussed the fit in respect of country demand for RBF, including the potential of the "double-bind" where partner governments may feel constrained by Bank requirements not only on inputs (e.g. on procurement regulations) but also outputs too. There may in fact be cases, such as those in which there is an alignment of interests and incentives between a national and state level government, where the discipline of RBF is beneficial for all concerned. John Newman pointed to this scenario in his paper, including on how he has seen it unfold in Brazil. Thomas O'Brien argued that the same conditions applied in a recently approved major Bank support for rural roads in India (known as the PMGSY program).

The applicability of RBF also is inextricably linked to measurement issues. John Newman brought this out very clearly in his discussion of malnutrition. There he showed that innovative work from the South Asia Food and Nutrition Security Initiative (SAFSNI) is defining the causality chain in reducing malnutrition, and putting in place careful measurements to produce data which can be analyzed and interrogated rigorously. This more systematic approach holds out the promise of making bigger and faster inroads in the fight against malnutrition, potentially avoiding huge numbers of child fatalities over coming years. Ravi Kanbur also emphasized the importance of solid and accurate measurement of results, and noted that for SAR, a region-wide focus on statistical systems is an essential component, indeed a hallmark, of a results-based focus. SAR is working on this topic as illustrated, for example, in an innovative policy-based support of $100m to the Government of India to strengthen its approach to statistics which help tackle poverty.

Monitoring and evaluation

Good monitoring and evaluation (M&E) is something that many adhere or aspire too, but perhaps not so many achieve in practice. Alan Gelb pointed out that according to the Independent Evaluation Group (IEG) of the World Bank, the average quality of M&E in Bank supported investments is low. Further he noted that the incentives and feedback loop for good M&E are weak in many cases, both within and outside the Bank. Many speakers, including Thomas O'Brien, argued that a drive for results-based financing would be a huge fillip for better M&E. Why? Because M&E is central to the approach and will be subject to much greater scrutiny since it is needed to unlock disbursements within this instrument. The good news is that the capacity and strength of M&E systems can be built over time, and John Newman's description of how this was done very carefully and explicitly in Minas Gerais was instructive.

Objections to RBF

Proponents of RBF must be able to respond to legitimate concerns, and several were discussed during the proceedings.

The "cash-flow" problem and unpredictability of the flow of aid

Some argue that a Bank loan is often required to pay for the investments which then deliver the results – and paying for results is putting the cart before the horse. Without the money up-front, how can results be achieved? There are several ways of dealing with this. One is to note that for many of the Bank's clients, while Bank finance is important it is not necessarily critical and nor is it by any means the only resource available. So in many cases countries will have their own resources which can pump prime an initiative which then unlocks Bank funding in due course. Another option can be to calibrate results on an ascending scale, and actually release a first slice of Bank funding at the outset of a project (even though first-phase results are likely to be modest) and so providing resources to deliver more results later. Thomas O'Brien pointed out that in SAR's experience in Pakistan, the RBF operations have in fact disbursed finance to the client faster and at bigger scale than in a traditional investment operation where specific and limited inputs are financed directly.

The lack of fit of RBF

There are topics or situations in which it is not clear that RBF fits. For example, a project which aims at institutional strengthening, or policy reform, may not have obvious and clearly measureable results. A project which is predicated on building a specific piece of infrastructure, such as a large bridge, may indeed want to place a great deal of emphasis on the design and conduct of the construction and maintenance, well before the long term results of the bridge opening become apparent. This point is well taken and most of the speakers were at pains to point out that RBF is not a catch-all, and should be used in a targeted way where the fit is snug.

Maintaining standards and safeguards

Ravi Kanbur emphasized the need for a careful dialogue with Governments to explore those situations in which the RBF approach is likely to work best, and for which they are comfortable with the standards and safeguards that the Bank expects to be deployed with its financing. Thomas O'Brien pointed out that the long term benefits and impact of sound fiduciary and safeguards practices are best delivered through country-based regulations and practices. Much of the practice to date has not really attained that goal, but rather has delivered on robust fiduciary, environmental, and social standards through "ring fencing" within a specific development finance investment (from the Bank or any other institution). The challenge with RBF which supports large scale (often national) programs is to work with country capacity so the benefits of solid safeguard approaches are delivered across all sources of finance (including domestic spending which typically greatly exceeds that coming from official international sources). As the World Bank develops its approach, the new ways in which appropriate fiduciary and others standards are to be met will have to be delineated.

3. Next Steps

Many of our partner countries in the South Asia Region have expressed a keen interest in and demand for RBF approaches, and so in response we intend to fully explore this path. There is no doubt that in the right circumstances this approach can produce more effective inroads in the fight against poverty. In doing this we will need to give thought and energy to several items:

- Develop a combination of in-house and external expertise and technical knowledge associated with results-based financing, tailoring it to new areas in which this approach can be applied
- Set a broad direction for the active use of results-based financing approach, and the criteria for judging whether a new investment is a good fit
- Work closely with countries to explore this avenue with them, and respond to their needs and aspirations
- Engage with colleagues elsewhere across the Bank and with other development partners, to help develop and then implement new approaches to RBF
- Reach out across the broad development community to gather ideas and feedback which helps refine and improve the way in which this approach can be deployed in the future

We hope that this compendium is a useful contribution to advancing the discussion on results based development assistance in South Asia and beyond.

2. Framing a Results-Based Approach to Development Assistance

Ravi Kanbur, T.H. Lee Professor of World Affairs, International Professor of Applied Economics, Professor of Economics, Cornell University

1. Introduction

Results based development assistance is in vogue, again (if it ever went away). The rhetoric certainly ramped up with the IDA-16 negotiations, and the surge of randomized controlled trials in development projects of NGOs and some donor agencies has contributed to the heat. What if anything is new in all this? And what does it mean for the World Bank, especially at the sharp end of operations?

A simple way of framing the discussion is to return to the old distinction between inputs, outputs and outcomes. In fact, as becomes clear the moment we try to actually apply the categorization, this is a continuum, with any point along it being an outcome of a process that went before it, and an input to the next step in the chain. Teachers' pay is an input into teaching hours which is an input into teacher/pupil ratios, which is an input into learning and skill acquisition (which can be measured by test scores), which in turn is an input into earnings (if the labor market does indeed reward these skills). Finance for building roads is an input into the actual roads of a certain standard, which are an input into improved transportation, which is in turn an input into improved economic activity (connection to markets, which is an input into generation of higher incomes) and social outcomes (better access to hospitals, which is an input into lower maternal mortality). Reduction of tariffs is an input into better reflection of comparative advantage, which is an input into more efficient allocation of resources, which is an input into greater income and also greater productivity growth, which is an input into higher long term growth. And so on.

Thus a "results based approach" has to specify which segment of the continuum it is referring to—what are the outcomes being looked for and what are the inputs being considered? But even if this is done, there remain the problems of measurement and of attribution. The measurement problem is not easy but can be addressed in principle. The attribution problem is also not easy, and it can also be addressed in principle, but it presents huge problems in concept and in practice. These three issues, (i) specifying the inputs and the outcomes to flow from them, (ii) measuring the inputs and outcomes, and (iii) attributing outcomes to inputs, are of course interconnected and intertwined. Without measurement, attribution is not possible, neither is specification of outcome targets. An initial specification of the inputs-to- outcomes link requires there to be some model of attribution, if only implicit. And which segment of the continuum is chosen for focus will affect the difficulties of measurement and attribution. Nevertheless, I believe that these three broad headings can help to frame and then to sharpen the discourse on results based development assistance.

2. Specifying Inputs and Outcomes

It is of course obvious that any proposal to use scarce public resources, in this case development assistance, has to specify the use of the resources and the outcomes anticipated. Indeed, that is what all project proposals do, by definition. The issues, however, are (i) how far along the continuum are the final outcomes, and (ii) how well established is the causal link between the inputs (which is what the assistance will finance) and the outcomes? Now, the problematic is that the further along the continuum is the outcome, i.e. the closer it is to some final outcome that can be clearly recognized as a goal of development (for example, an element of the MDGs), the less likely it is that a strong causal link can be argued between the inputs of the project and that outcome. At least, that is a question that can be asked of the project proposal. In many cases there is a tradeoff. The shorter the chain from input to the outcome the project is claiming, the more easily is the causal chain established. But then, most likely, the further will be this outcome from a final outcome of development interest.

Let us consider some examples. A roads project proposes inputs of finance and Bank staff time and an outcome of roads of such-and-such length and such-and-such quality. This chain is not without its own problems and issues, but it has the virtue of being quite short, and the attendant virtue of causality being relatively easy to establish. Indeed, whether these outcomes actually occurred is hopefully the sort of thing that IEG evaluations would investigate. But roads are not an end in themselves. The reduction of maternal mortality is an agreed end in the MDGs, and roads to hospitals are recognized as an important element I this reduction. If a roads project proposed this as an outcome it should receive high marks for its ambition but, equally, close scrutiny on the causal chain from project inputs to this outcome in the specific context under consideration.

Some Analytical and Advisory (AAA) work proposes inputs of Bank staff and consultant time for a country economic report, or economic report on a state in the country. The specific monitorable output is the report itself, plus dissemination workshops, etc. This short chain is easy to link together causally, but a report is not an end in itself—thousands of Bank reports are gathering dust on hundreds of shelves. When pressed, and they should be so pressed, presumably the project team would say that the report would have a catalytic effect on debate and discourse in the economy or in the state or in the sector. If they do say this, then they would have to specify how and how it would be measured (more on measurement in the next section).

A Development Policy Loan (DPL) proposes reform of power pricing. The inputs are Bank staff time and financial resources, which will provide budgetary support as the reforms are undertaken. Here the immediate outcomes are the reforms, which can be monitored and are likely to be part of the conditions. But pricing reforms are not an end in themselves. They are predicated rather on a strong belief in a causal model that links these reforms to better outcomes on efficiency of resource allocation, leading to higher growth and thus, eventually, to poverty reduction, which indeed has claim to be a final development goal. But here we will run into problems of attribution in practice, even if we buy the causal link in theory (more on causality later).

The overall point is that if a results based focus is to mean anything, then all projects (investment, AAA or DPL) must specify outcome, how it will be measured, and how the outcome will be attributed to the project inputs. It is recognized that this is not easy and there is no perfect answer. But an urging or an even a requirement, from the review process, to strive for an answer is an essential component of a results based framework.

3. Measurement

A results based focus also requires that each proposal for development assistance not only specify outcomes, but how they are to be measured and monitored. In the previous discussion on inputs and outcomes I highlighted the inputs-outcomes distance on the one hand, and project outcomes-final development outcomes distance on the other. This distinction is not so relevant here. Outcomes can be measured all along the continuum from "pure inputs" at one end to "pure development outcomes" at the other. The real issue is that measurement is more difficult for some outputs and outcomes than for others. And, in some cases, qualitative techniques may have to complement or even replace quantitative ones.

Thus, for example, at the input end of the spectrum, reports and workshops are easy to enumerate and quantify. But judging their quality is, well, a qualitative exercise involving a fair amount of judgment. Thus when an AAA proposal says it will produce so many reports and do so many dissemination workshops, it is legitimate to ask how the quality of this will be assessed and reviewed. Of course, the Bank has peer reviewers from within and outside the Bank. But several issues arise. The first is the independence of peer reviewers who are, in effect, being paid by the project team to review their product. The answer to this cannot be found with the project team, but with the constitution of a separate funding mechanism for peer review outside the normal process of project preparation—perhaps at the Vice President (VP) level. Second is that peer reviewers typically review a concrete product like the report, rather than, for example, the impact of the dissemination workshops. Again, IEG evaluations are meant to address this, but we are discussing here mechanisms within the Vice Presidency.

Consider now the roads project that goes beyond claiming as output roads of certain length and quality (mechanisms for monitoring which can be specified relatively easily). Suppose it claims an improvement in market related transport or easier access to hospitals. For such claims to be measured there are two requirements. First, there has to be baseline data on these dimensions before the project. A project proposal without baseline data, on magnitudes it claims to affect as a project outcome, cannot be given high marks in a results based focus. Second, given the baseline, the project must have, as an integral part of project design, a mechanism for collecting and analyzing data after the project is completed. The issue of attribution will be taken up in the next section, but without such provision, there cannot be a results based focus.

Finally, let us discuss some issues that arise from DPLs that purport to have sector wide or economy wide impacts through policy change. Again, the issue of attribution is a difficult one that will be taken up presently, but if DPLs make claims on economy wide impact on development prospects then this requires us to have economy wide data on key development outcomes. This is not necessarily the responsibility of any particular DPL, but a results based focus for a Vice Presidency would mean systematic support for collection, maintenance and dissemination of sector-wide, state-wide, and economy-wide data on key development outcomes, certainly for MDG outcomes. A VP-wide focus on statistical systems seems to me to be an essential complement, indeed a hallmark, of a results based focus.

4. Causality and Attribution

If development assistance is to be judged by results, and yet all that development assistance can provide are inputs, then causality from inputs to outcomes, and attribution of outcomes to inputs, must be a core concern of a results based approach. Causality and attribution requires specification of which part of the continuum is being assessed, and it requires measurement of inputs and outcomes when those are identified along the continuum. But even with these in hand, there are other multifold problems. Some are amenable to solution, others less so.

To begin at the beginning, every proposal for use of development assistance presumably has some causal mechanism, hopefully explicit but quite often implicit, whereby the inputs of the project are argued to influence development outcomes. If the inputs are quite close in the continuum to the development outcome in question, for example, oral rehydration therapy (ORT) as an input to preventing child mortality, then the causal chain is shorter and easier to establish. In this case, in any event, there is considerable global evidence, and global consensus based on that evidence, on the efficacy of ORT. Here the issue would be how the inputs of the project (for example, training mothers to do ORT) would lead to better ORT, the causal link from ORT to reduced child mortality being accepted.

However, things are not always so easy. The farther is the input from the outcome of interest, the more difficult is the establishment of causality, and the less likely is there to be an established global consensus on the causal chain from the inputs of the project to the development outcome. Sometimes such a consensus does not exist even when input and outcome are close to each other, because of the many different factors at play, and the difficulty of attributing causality to the input the project is emphasized. For example, it is now well established that if complications set in during child birth then there is a very limited time in which to get the mother to a hospital for treatment in order to prevent maternal mortality. But preventing maternal mortality depends on a number of factors—the road, of course, but also local knowledge and sensibility to take the woman to hospital, timely availability of transportation, and availability of medical personnel and drugs at the hospital. Of course all of these things matter, and having better roads should reduce mortality. But by how much? Comparing before and after the road, or with and without the road, how much of the difference in maternal mortality can truly be attributed to better roads, especially if some of the factors that influence the difference are not measured or measurable?

The problems are particularly acute when the distance between input and outcome is large—for example from economy wide reforms to economy wide performance and thence to poverty reduction and human development progress. Much of the received wisdom in this area comes from the accretion of results from cross-country regression analysis. But there is a vigorous debate on these findings, in their own terms and in terms of their implications for country specific policies. The problem is particularly acute when a country is an outlier on the cross-country regression line. For example, Bangladesh's performance on poverty reduction and other social indicators is far superior relative to its score on transparency and corruption. This is an element of the "Bangladesh paradox" which has been much discussed in the literature. More generally, translating cross-country associations into country specific causal links between inputs and outcomes remains problematic.

A results based approach to development assistance will therefore interrogate closely the presumed causal links between inputs and outcomes in any proposed use of funds, asking whether the evidence really does support the links underpinning the project. However, there is another type of focus that a results based approach can bring, namely to use the project itself to test the hypothesized causal links. This takes us back to the issue of measurement, for without appropriate measurement—with a baseline and then data on outcomes—no causal analysis can be done in any case. But it also takes us further, to the issue of whether a causal assessment can be built in as part of the project itself. This will of course contribute to the evaluation of the project itself, but it can also contribute to other projects as understanding on this causal link is firmed up. There is clearly an externality here. The project level incentives will tend to lead to underinvestment in this dimension, so there is an argument for a central fund to support such evaluation and assessment.

This is the point at which we can also address the current debate on randomized controlled trials (RCTs). There is no question that they are a major innovation in evaluation and should be considered as an integral part of a results based focus. However, this integration should be done with several points in mind. First, not all types of development assistance projects can be evaluated with RCTs. A road project, for example, cannot be randomly assigned—although the idea of having a control for comparison (an area that is similar to the project area but without the project) is important in evaluation. Second, economy wide or sometimes sector wide programs of policy reform cannot be evaluated in the RCT framework—in fact, in this case not only is randomization not possible, neither is control. Third, even when the conditions for RCTs are right, the results learnt have to be interpreted carefully to draw broader lessons. The basic issue is the tension between "internal validity" and external validity". RCTs are strong on the former, because they control for all sources of variation other than the project in question, but for this very reason they may be weak on the latter if we apply the findings to very different situations. Thus, for example, a careful RCT on whether incentives for better teacher attendance improves children's test scores in a particular village in Bangladesh may or may not tell us (and certainly not in terms of quantitative magnitude) as much about the same intervention in schools in India, or even in other parts of Bangladesh. None of this is to deny the importance of the RCT approach, and indeed the RCT "mindset" for a results based approach. It is simply to warn that it is not a panacea, nor the silver bullet.

5. Results Based Financing

All of the above asks what a results based approach would imply within the current Bank framework, where operations are prepared and approved with a predicted link between input and outcome, and then evaluated afterwards. The discussion has emphasized how the link between inputs and outcomes can be better interrogated during project preparation and better evaluated after the project. (There is also the issue of project supervision, which can also be an important part of a results focus, but about which this note has not said anything.)

However, there is a more radical direction that a results based approach can take, and the direction comes from the uncertainties of knowledge in linking inputs to outcomes and the requirement of a project that the inputs be definite and specified. It accepts that the development agency does not have sufficient ground level knowledge, and cannot engineer ground level commitment, to ensure that outcomes flow from inputs. It can give advice, of course, but not determine the inputs through an assumed input-outcomes linkage. The logical conclusion to this line of argument is simply to make development financing available upon delivery of agreed upon results. It has also been called "Cash on Delivery" aid. If a country can deliver an improvement in maternal mortality, say, then why should it matter what particular combination of inputs it used to deliver this outcome? So long as there is delivery, cash should flow.

In one sense this approach is standard. The recipient does something before receiving the assistance. In investment projects, the roads are built and expenditure claims are sent in. In development policy loans, the tranche is released when tranche conditions are met. In the case of results based financing (RBF) or cash on delivery (COD), the conditionality is straight forwardly the achievement of agreed results—it is just that these results are closer to the outcomes end than the input end of the input-outcomes continuum.

Of course, in another sense this approach is very different indeed from standard World Bank investment projects. The differences can be highlighted by posing the following questions. First, how do we know how much to pay for what outcomes? In standard projects, we paid for inputs and these inputs were costed and that is how we knew the total project amount. Of course, since the whole object was to get outcomes, and the outcomes were (presumably) spelled out, we are paying for those outcomes, but only indirectly and with no guarantee of those outcomes. Now we are paying for outcomes after they occur, without specifying in any way the inputs. This should highlight the issue of much to pay for what outcomes. If the old style project document did indeed specify outcomes, then one benchmark is to offer the project cost calculation as COD instead. But this leads to the second question, how can this be fitted into an investment project framework, with its procurement guidelines and procedures? The answer is that it cannot easily be so fitted, and we might consider instead viewing COD projects as program support, with payment into the budget. A third question is, what about upfront costs to achieve the outcomes? One answer to this is that if the government wants the outcomes badly enough it will put up its own resources in anticipation of being repaid when the outcome is achieved. But another answer is that perhaps some can be paid up front, with payment on other COD projects nearing completion as the guarantee that the government will not abandon the project once it receives the first tranche.

In developing a results based approach to financing, a particular danger must be avoided. The basic tradeoff is between conditioning finance on outcomes and conditioning them on inputs. In a conventional investment project finance is conditioned completely on inputs. Eligible expenditures are specified, and procurement processes for inputs of the projects are also specified. In addition, investment projects have to comply with environmental and other safeguards. Adding conditions on outcomes to this structure would be over determination and overburdening for the country. In effect, the quid pro quo for making outcomes a condition of disbursal is to loosen the conditions on inputs. There is of course considerable debate currently at all levels in the Bank on these issues, in the context of developing new investment lending instruments.

Clearly, many questions need to be asked and answered about COD or RBF in the context of Bank assistance. But the case for the Bank doing pilots, perhaps financed by trust funds, is strong in the framework of a results focus.

6. Implications for the South Asia Vice Presidency

The discussion has provided a framework for structuring discussion about implementing a results focus to Bank operations in a Vice Presidency such as South Asia. Within the context of standard Bank instruments the framework suggests (i) greater clarity in project documents on outcomes, and on evidence for the causal link from project inputs to outcomes, (ii) an emphasis on measurement, with baseline data and data on outcomes during and after the project, and (iii) a focus on using the project itself to test the hypothesized causality between inputs and outcomes, through a range of methods including but not restricted to randomized controlled trials. Outside the Bank's conventional instruments, the discussion suggests a major push towards operations with results based financing, or "cash on delivery".

The points made in this note will probably be accepted by most South Asia (indeed Bank) staff, and they will probably say that they are already doing much of this. And yet in practice we still see a predominance of the "input focused" mindset. This is true in investment lending and in development policy loans. The introduction of new instruments, such as the results-based investment loans which are being discussed currently, will help, but more is needed to overcome the input focused mindset and move to an outcomes based perspective. Essentially, the push needs to come from the top. Consider the following objective: over the next three years the South Asia Vice Presidency will move to have the Bank's largest results-based-financing portfolio, defined by the following illustrative criteria: baseline data, systems for collecting information on outcomes, benchmarking of performance with respect to outcomes, disbursement linked to outcome performance, and flexibility on specific inputs used to deliver outcomes. Such an objective has the potential to transform the South Asia portfolio, by giving countries more control over their development processes and judging the partnership by actual development outcomes.

3. Framing Results: Motivation and Incentives

Alan Gelb, Senior Fellow, Center for Global Development

I would like to make a few comments on the excellent points raised by Ravi and then focus on a topic that he does not emphasize: the question of motivation and incentives.

First, especially in the early stage – I think that it is well not to be over-ambitious in always seeking to stretch the indicator for disbursements to final outcomes. Certainly we should make efforts to go down the results chain. But, to take an example, to assess a road to the hospital by declines in the infant mortality rate is stretching credibility, given the range of other factors that will influence mortality. Clarity, credibility and simplicity are desirable elements of the results chain.

Second, new technology is making it easier to specify and measure a range of important outcomes even if these are not the MDGs or macro goals like the level of private investment. For roads, for example, we can measure traffic flow and transit times, to see if there is really an improvement. This is better than just measuring miles rehabilitated; the problem, as in much of West Africa, may be road-blocks. For ports, we can measure container dwell times, for customs and border posts, clearance and transit times. For power, we can measure access and quality of service in great detail. We can use remote reporting via cell phones to monitor whether water standpipes across the country are in working order or not. Some countries are building such feedback directly into points of service. For example, I recently visited a country where one punched a button immediately after the agent had completed the passport check. There were three choices: smiley face, neutral, and frown, and I am sure the time was recorded also. What a great way to monitor the quality of service delivery by immigration!

Some years ago Vijaya Ramachandran, a colleague at the Center for Global Development (CGD) and I proposed issuing “service guarantees”. As part of a major infrastructure project – say to modernize power supply or the port and customs -- firms would be able to purchase insurance against poor power delivery or slow clearance. To prevent manipulation, this would be on average performance, not on the experience of the particular firm. The insurance would be underwritten by an aid-funded guarantee. The purpose of this facility is not really to compensate the firms for losses, though it would help. It is to create a mechanism whereby poor performance, in the form of service failures, would immediately be brought to the high-level attention of both governments and donors. To put the implications in the most basic form – funds that could otherwise be used to support education or health are being paid to firms because IDA’s ports project is failing to perform. This would be noticed! The idea was discussed with the World Bank’s Multilateral; Investment guarantee Agency (MIGA) and they showed some interest but nothing has moved yet.

Third, we should note that the EU is beginning a review of their partly results-based country budget support instrument. It involves a large fixed tranche plus several performance-based tranches. Not all are conditional on true outcomes but some are: on average about 70% of the variable tranches have been disbursed. The last assessment was in 2005. Work began on this approach in 1996; in thinking about the integration of results into funding decisions, the EU has been ahead of the Bank.

Next, on the topic of incentives, who actually has the incentives to condition financing on results? It is not clear, either on the donor side or on the clients' side.

I recently produced a CGD Working Paper on this (WP 227), arguing that several things needed to be done. First, incentives need to be rebalanced. Procurement and safeguards have short feedback loops. A request for an Inspection Panel review can be submitted even before the project is approved by the Board (as was the case with the South Africa coal power project) and the feedback-outcome loop will take only several months. In contrast, IEG evaluations of project performance come years later. A staff member 54 or over who takes a typical project to the Board will be retired before the IEG evaluation! The paper proposes a short feedback loop for performance, comparable to safeguards, focusing on project design, data and M&E system, with external independent assessment and all fully public. IEG may be reluctant to do this, but it could manage an independent process. Or this role could be that of the successor to the Bank's Quality Assurance Group (QAG). Full transparency could substitute for mandatory assessment – it would be difficult to explain (publicly of course and for the record) a decision to proceed without external scrutiny.

This process would force better data arrangements and M&E. Many believe that good M&E will only uncover problems and is therefore resisted. But we know from IEG assessments that the projects with the best M&E also typically are rated as more successful – and this is not driven by country characteristics. The result is reassuring, and we need to get the word out. The process would also offer more recognition to task managers who are able to design and implement good projects in bad conditions.

Second, we also need to increase the link between results and financing in terms of the country envelope (this refers to IDA countries). Ravi has urged complementing the Bank's Country Policy and Institutional Assessment (CPIA) by development outcomes. I will go in a slightly different direction to focus on aid effectiveness, and with a special focus on Fragile States. In the early days of IDA, indicative allocations were based on three equal criteria: country performance on growth, country performance on equity and the Bank's impact on policy and projects (in a 1977 Technical Note by Mahbub ul Haque). Today, the effective weight of the portfolio in IDA allocations is negligible. A 10% increase in the CPIA at midpoint increases a country allocation by 55.3%. A 10% increase in the portfolio rating increases allocation by only 4%. For IDA countries (those that are not capped) and programs seeking to increase resources there is no incentive to have well performing projects. It is better to contest the CPIA especially block 4 on governance which has most of the weight.

The current portfolio measure may not, of course, be the right way to introduce project performance, especially for Fragile States where we should expect many poorly performing projects as well as some good ones. We need to be able to encourage what is working and use the base allocation to see what works. As far as possible projects should be designed in a scale-up mode, with independent (and rigorous) assessment at the mid-term, which is typically three years after the start of implementation. We would reduce the automatic or base allocations to Fragile States but allow them a long leash if they are able to produce well-performing projects able to produce results that could be scaled up.

This change would transform the dialog between the Bank and Fragile States. It would challenge them to work with IDA to demonstrate effectiveness and reward them if they did so; if they ended up with little it would be clear to all parties that the reason was not simply a judgment on their governance and economic management, but also a response to the failure to use resources well. It would also increase incentives to develop more results-based financing mechanisms, including the new innovative instruments that the Bank is working on and other innovations along the lines of Cash on Delivery (COD) aid.

As Fragile States are likely to become an ever-larger share of IDA's portfolio a consideration of these issues is becoming more urgent.

Finally, I would like to stress the importance of monitoring the introduction of new results based financing instruments and its impact on a real-time basis, not for public relations purposes; but because clear and independent documentation of achievements will be needed to counter the (hopefully few) criticisms and odd scandals that are bound to ensue, whether in country procurement or safeguards systems. The more one can achieve results and demonstrate them, the more one will be able to survive occasional problems with the use of the instrument. This could be a good area to work collaboratively with an outside organization.

4. Results-Based Operations at the World Bank

Ariel Fiszbein, Chief Economist, Human Development Network, The World Bank

All World Bank operations are designed and formulated to support the achievement of results –as expressed in the Project Development Objectives (PDOs) and the associated results framework. Different instruments pursue results following different approaches depending on the nature of the bottlenecks and constraints the country faces to achieve those results.

Development Policy Operations (DPLs) are focused on discrete policy actions within the direct control of governments and link disbursements to evidence that those actions have been adopted. DPOs are a practical and effective way of supporting the achievement of results when the policy actions are critical in creating the enabling conditions for improving results. This is the case, for example, when new regulations are required for the better functioning of markets or new policy frameworks are necessary to improve government efficiency. Typically these actions do not imply direct expenditures. By focusing on the definition and monitoring of the appropriate policy framework, DPOs put the emphasis of the Bank-country relation on what is understood to be the most critical step to enable results.

Specific Investment Operations (SILs) are a heterogeneous category. But their common characteristic is the financing of specific government investment activities, which involve a set of transactions for the purchase of goods and services. SILs have been proven to be a practical and effective way of supporting the achievement of results when close control over inputs (e.g. construction, technology, etc.) is essential to the achievement of results. This is typically the case when technical design and implementation challenges are critical bottlenecks. Without them results cannot be achieved –and to the extent technical challenges (at both the design and implementation phase) are addressed the probability that results will be achieved is very high.

Often such situations involve discrete, one-off, activities (for example the construction of a large infrastructure project or the purchase of expensive and technically sensitive equipment). By focusing on the proper implementation of individual transactions, SILs put the emphasis of the Bank-client relation on making sure that the right inputs and technology are in place.

Many of the development challenges faced by client countries, however, cannot be addressed just through discrete policy actions or through the proper technical implementation of a project. For example, improving service delivery (e.g. better maintained roads, functioning schools and health clinics, effective agricultural extension services) may well require policy actions (e.g. a decentralization law) and will typically involve some discrete investment activity (e.g. constructing new schools or purchasing new equipment for the road maintenance agency). But these are seldom sufficient for the achievement of results: schools can be built but teachers may remain absent; health clinics may have new equipment but essential drugs may not be available at the point of service;

rural roads may remain un-maintained in spite of the existence of idle equipment. Addressing such bottlenecks involves changes in management practices and behaviors by service providers and users alike. Improved policy frameworks may be necessary and more or better inputs may be essential –but without stronger incentives and accountability mechanisms improved policy frameworks and inputs are often unlikely to yield results. Typically the relevant activities are not discrete or one-off but recurrent and sustained over time.

Over the years World Bank teams and their counterparts have found that the intensive focus on inputs implicit in the transaction-driven approach of traditional SILs is not conducive to achievement of results when bottlenecks are of a behavioral nature and addressing them require changes in practices. They have found that putting the emphasis of the Bank-client relationship on the monitoring of results is essential and that linking Bank disbursements to the achievement of well specified results that materialize over time provides a much stronger set of incentives than the strong focus on input controls implicit in traditional SILs. While controlling inputs is effective when challenges are mostly discrete (and of a technical nature), it can be stifling when challenges are recurrent (and of a behavioral nature).

Unlike in the case of the more traditional 'investment project' –where boundaries between what is being financed and what is outside the purview of the project are more or less clear— focusing attention on service delivery or, more generally, on improving the performance of government programs makes it complicated to establish neat boundaries between the 'Bank-financed project' and 'Government-financed programs'. Moreover, moving away from an input-control approach to a result-monitoring one implies a stronger reliance on, and integration with, the Government's own systems and procedures.

The previous discussion highlights the fact that the focus of a Results Based Operation (RBO) is to support a 'well-defined' government program. A critical part of the Bank's appraisal process ought to be an assessment of the strategic relevance and quality of that program. In simple terms, this means assessing whether the program is 'good enough' to justify the Bank's support in the form of a RBO.

In that regard it is important to clarify upfront a few important considerations:

- A RBO is not necessarily the right instrument under all circumstances. It presumes that the *essential* policy framework enabling a results-based program is in place and the client can benefit from Bank support in implementing that framework. It also presumes that the implementation challenges are to a great extent institutional (as opposed to strictly technical) and demand a focus on incentives and accountability for results. In contrast with DPLs, RBOs allow the Bank to accompany the government through the implementation process and contribute technically as well as financially. A DPL remains a stronger instrument when the bottleneck is the adoption of a policy framework. And when the key bottleneck is technical, a focus on managing inputs (as in the traditional SIL) would still be needed and desirable.

- The Bank must be realistic and pragmatic in its assessment of what is 'good enough' when considering whether to adopt a RBO to support a government program. This operates at two levels. First, in terms of scope. A government program can be as large and complex as an entire sector or sub-sector, or small and simple as in the case of a specific intervention in a particular geographical area. Quite often it will be a large but 'contained' flagship initiative. Second, in terms of the benchmark used to assess the program's quality –and the associated risks. A RBO operation should not be seen as an instrument to support only the most sophisticated clients but as an instrument to support institutional development, including in low income and low capacity settings.

The starting point to define the desirability of the Bank's support is an assessment of the strategic relevance of the government program. Is it addressing an important goal in the sector? Can we comfortably claim that if it succeeds sector performance would have improved in a meaningful way? This will typically require some, however rough, element of benchmarking. For example, if the stated goal of the government program is to improve the quality of education the assessment ought to provide some indication that low quality is indeed a serious problem. Or if the goal is to support an expansion in 'pro-poor access' to water connections, it should provide some evidence that the country (or region) is lacking in this area.

An assessment of the soundness of the program is a second important aspect to be considered. Is the approach proposed by the government consistent with lessons and evidence from past experience in the country and abroad? Is it addressing the key 'behavioral' bottlenecks underlying the poor performance at baseline the program is seeking to correct? For example, if the problem is low access to water connections among low income households due to high connection costs is the program proposing a reasonable strategy to increase effective demand for water connections? Or if the problem is low quality of education services, has the program identified the key drivers for change as they relate to availability of pedagogical inputs or teacher presence and motivation?

The assessment of technical soundness must be complemented with an assessment of institutional arrangements. The premise behind the adoption of a RBO approach is that the implementation of the program is feasible and, under reasonable circumstances, will not be derailed by weak systems (e.g. monitoring and evaluation) or opposition and/or lack of interest on the part of key stakeholders. Such an assessment should allow the identification of risks and implementation bottlenecks, as well as the necessary plans for implementation support.

It can be argued that similar considerations apply to other operational approaches. While true, there is a strong sense that because of their nature, RBOs demand a more consistent effort to assess the government's program in this broader context. In that sense, RBOs may present a unique opportunity for the Bank to continue enhancing its ongoing efforts to mainstream governance and accountability issues in operational work, including by strengthening work and collaboration among sector units and partnerships with client agencies.

The key aspect of a RBO is that it finances outputs/outcomes, instead of inputs. This requires moving away from traditional costing exercises (which focus on inputs) and focus on assessing the level, efficiency and equity of the expenditures that form part of the government's program to be financed. An essential aspect of this assessment is that it focuses on the entire government program (however broadly or narrowly it is defined) and not only on the share or components financed by the Bank. This may increase the stakes in terms of the expenditure analysis that is required depending on the size and complexity of the government program (not the size of the Bank's financing). In other words, the Bank's appraisal must consider the efficiency, equity and sustainability of the entire government program. The extent of such an analysis will depend on the program in question. If, for example, the Bank is to support a whole sector program or a large flagship program, it will be important to ensure that they are part of a medium term expenditure framework.[2] Similarly, it is expected that issues of equity and efficiency would have been identified as part of the overall strategic framework and would guide the design of the program itself. The economic analysis should, equally, focus on the program rather than on the share financed by the Bank operation.

Finally, the importance of having solid results frameworks and plans for monitoring and evaluation has long been recognized as a critical aspect in all operations –and, often, an area of weakness in many operations. RBOs, by making results the trigger for disbursements (as opposed to inputs which trigger disbursements in traditional investment lending operations) represent a fundamental change in this regard: measuring results is not just one more aspect of the project but the factor that drives its implementation. In other words, RBOs operations have strong built-in incentives for both the client and the Bank to focus attention on results measurement.

The Bank has been implementing RBOs for a number of years under the existing policy framework for investment lending. In particular, within the Human Development sectors results-based approaches have become very popular and are in high demand by staff –and increasingly by clients.

Consultations with operational staff indicate, however, that there remain many serious bottlenecks to the use of RBOs in any systematic way. The most important one is the lack of a clear and transparent policy framework under which RBOs can be designed and implemented. Using the existing investment lending framework creates too many complications, adds transactions costs and gives too much latitude to the views and preferences of individual managers and groups within the various regions on whether and how such approaches can be pursued.

The Bank is engaged in a process of developing a proposal for a new type of lending instrument which would address these concerns. New operational policies will need to be developed to make this new instrument a reality. At the same time, a number of managerial challenges will need to be addressed as well.

[2] However, it should not be assumed that a full macro and fiscal analysis is needed as background to a RBO.

First, the type of analysis proposed above will demand an enhancement of upstream analytical work –typically conducted as ESW and not as part of project preparation. The challenge of such enhancement should not be minimized. In some cases this will require a change in the skills mix within operational units. But more significantly, it will most likely require a change in the way work program agreements between Country Management Units (CMUs) and Sector Management Units (SMUs) are defined –focusing resources on upstream diagnostic work that provides the foundation for RBOs.

Second, and critically important, will be the enhancing the capacity of the Bank to support the strengthening of country monitoring and evaluation systems. This is not just about being able to track in a credible way whether the results of the government program are being achieved but also essential aspect of the promise of improved development effectiveness that is driving the movement toward RBOs. A range of approaches will be required: from technical support to statistical organizations (e.g. to enhance the quality of household survey systems), to the use of social audits and other monitoring and evaluation efforts by civil society organizations and users of services. In many cases, other complementary reforms (e.g. improving transparency through freedom of information) may be required to create an enabling environment for monitoring and evaluation.

Over the last decade the Bank has gained substantial experience in this area and many Bank-funded projects are already moving in that direction. But the scaling up of RBOs that the changed policy environment should enable will demand more and stronger efforts in this area.

5. Experiences, challenges, and opportunities for Results-Based Financing

John Newman, Lead Poverty Specialist
Economic Policy and Poverty, South Asia Region,
The World Bank

In his note on "Framing a Results-Based Approach to Development Assistance" and the handout on "Observations on Results Based Financing", Ravi Kanbur has given us both some practical suggestions and a challenge to figure out what it would take to make the South Asia portfolio at the World Bank predominantly results-focused. In this note I would like to make some observations and points related to that challenge. These observations are heavily influenced by my experience working as part of a World Bank team in Minas Gerais with the government's State for Results program and from work trying to identify what is needed to meet target improvements in nutritional outcomes.

The first point is that it will be difficult for the World Bank's South Asia portfolio to become predominantly results-focused if the countries themselves do not move towards a greater results orientation. One reason for this is that generating better results is not just about making better technical choices in investment projects, but about getting incentives within the system to make those better choices, having the information needed to make more informed choices and having the will to want to work systematically to improve outcomes. It is difficult to advance in aspects of incentives, information and modes of working if there is not a champion in the government that wants this to happen. That champion needs to be in a position that sits above individual sectors. It could be a governor of a state or province, a minister of finance or of planning.

As incomes grow and the middle class expands in South Asian countries, the political demand for better performance of the public sector will inevitably grow. That demand is not yet at the level that exists in, say Brazil, but it will grow and the World Bank has the potential to help accelerate the progress.

Ravi makes the point that working at the level of states or provinces, in conjunction with the Ministry of Finance or Planning at the national level, is likely to be the most productive way for the World Bank to generate a portfolio that is predominantly results-focused. I strongly agree with this view. It is unlikely that there would be uniform interest in the countries in generating a greater results orientation, but there are certainly some states and provinces that have already manifest interest. Champions exist in the states. There would be potential for a productive alliance among the National Planning Commission, Ministry of Finance, World Bank and a selected state or province to work on improving performance.

The new Program for Results (P4R) lending instruments would provide an ideal instrument for the Planning Commission and Ministry of Finance to use as an incentive to induce better performance from the states, without having to set aside money in their budget for the contingent liability[3]. If the state does not reach the desired results, the loan would not be disbursed and the federal government would have paid very little for having provided that incentive to the states. If the state does reach the desired results, then the national government would probably be happy to pay.

If there is a need to actually finance the interventions necessary to generate results (as opposed to just trying to create the proper incentives), one option would be to combine a Results Based Financing (RBF) loan, also known in the Bank as a Programme for Results (P4R) loan, with either a traditional investment loan or a DPL. An alternative option would be to introduce a Results Buy-Down, an innovation in the financial architecture that has been used for polio eradication.[4] Under this arrangement, a country takes out a loan with a development agency such as the World Bank and, if performance targets are met, grant money is used to Buy-down the cost of the loan. The Buy-down could be structured to cover some, part or all of the interest and principal.

In a P4R lending instrument, what is at risk for the government is whether the funds would be disbursed. In a Results Buy-Down, what is at risk for the government is the ultimate cost of the funds. There would be a maximum cost of the loan, given by the rate if results were not met, but the loan could be cheaper if results are achieved. In some circumstances, countries may prefer risking the cost of the funds to the risk of not getting the funds at all.

A Results Buy Down represents a different way of blending grants and loans from what is currently being done. Currently, IDA loans are blends, where grant money pledged for IDA is used to provide lower interest rate loans. The cost of the loan is not conditioned on results, although achieving better results can result in a higher allocation of IDA funds to the IDA-eligible country.

Figure 1 illustrates how the Cash on Delivery, P4R and Results Buy-Down would fit within the existing spectrum of aid instruments. Going from left to right, the terms would be progressively more stringent. For the expanded set of options (including Cash on Delivery grants and Results Buy-Downs) to be offered, there would obviously have to be close coordination among those offering loans and those offering grants. To date, the allocation of grant money to a particular country does not appear to have been determined as the outcome of a strategic choice among the possible options in Figure 1. One would hope that, with the emphasis on donor harmonization that exists today, there would be more of a strategic view as to how best to combine the different instruments that would be offered to different types of countries. A key decision for the development partners would be to determine how best to allocate the scarce amount of grant resources across the different Aid Instruments, some of which are tied more closely to results than others.

[3] This is similar to the Cash On Delivery model, suggested by the Center for Global Development. See http://www.cgdev.org/section/initiatives/_active/codaid)

[4] See http://www.fininnov.org/img/pdf/19%20-IDA%20buydowns%20Nigeria.pdf for a presentation on the experience with the Nigerian case

Figure 1: An Expanded Spectrum of Aid Instruments

As it is likely that the World Bank would soon start to work with aid instruments such as P4R loans in South Asia, it is useful to review some recent experiences where the World Bank has actively collaborated with state governments to help strengthen a results culture. The experience of Minas Gerais, under its State for Results program, provides a useful reference point for the work that might be carried out in the near future with states and provinces in South Asia[5]. The context of this work is important. The big push on results came after the state successfully resolved some serious stabilization issues in the state. Buoyed by that success, the Governor turned his attention to the next big challenge – the low level of public sector performance and set about instituting a "Management Shock". The effort was comprehensive, as the government worked on 11 main results areas across all sectors. The government selected indicators, set targets, created performance agreements with incentive payments to groups and improved M&E systems. The effort was supported by two World Bank projects that covered multiple sectors and had disbursement linked indicators.

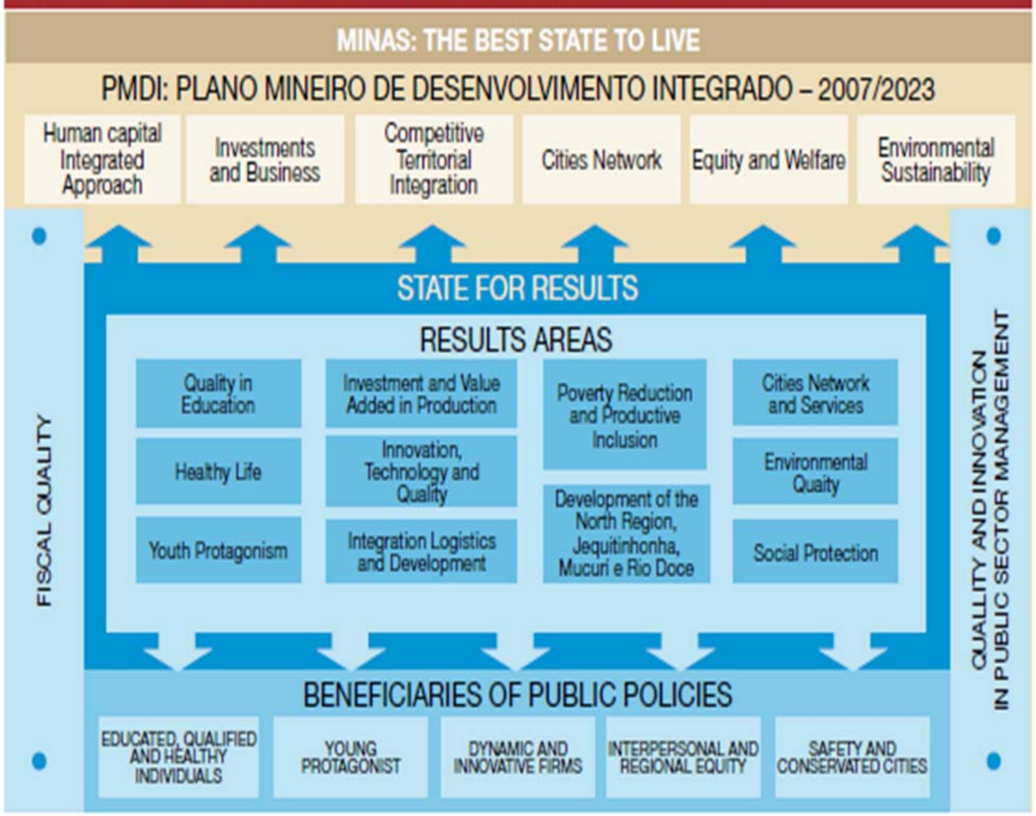
[5] More information on the experience in Minas Gerais can be found at http://siteresources.worldbank.org/EXTLACPROJECTSRESULTS/Resources/3177340-1288967051412/MinasPublicServices.pdf

Some of the key features of the approach are listed below and briefly discussed:

1. The government increased their attention on the **planning process** – both annual and medium-term terms and paid explicit attention to the relation between the two planning processes.

The task of managing results was facilitated by organizing the work of the state around 11 results areas, identified in **Figure 2** below. This helped manage the complexity. Working with hundreds of individual programs would have made it difficult to maintain high-level management attention on all programs. Each state or province in South Asia would have to define its own groupings, but it is likely that a similar process of aggregation would be required.

Figure 2 Government's Development Plan Strategic Map with Results Areas



2. The state introduced a practical and flexible approach to **Performance Contracts**, with features that should be replicated elsewhere.

At a first level of the Performance Contracts, the administration in Minas Gerais identified four separate areas for which agreements to reach targets would be sought: around final outcomes (resultados finalisticos), structural programs (projetos estruturadores), the sectoral agenda (agenda Setorial) and the rationalization of expenditure (racionalização do Gasto). These were set as part of a process of negotiation between the Secretary of Planning (the principal) and the sectoral ministries or institutions (the agents). A second level was also specified – where the head of the sectoral institution or ministry negotiated a set of targets with the lower level or more geographically disparate agencies that made up the sectoral ministry.

This approach was a little different from a strict contracting out model between a principal and agent, where the principal specifies what they would like to achieve in the final outcome and leaves it completely up to the agent to determine how best to achieve the outcome. By also including outputs and structural reforms (agenda setorial), the principal (in this case in Minas, the State for Results) was putting some weight on activities that they think go into the production process for generating the outputs, not just on the final outcomes. The relative weight attached to the outcomes and the outputs varied across sectors. Where there was more data, a longer history of working with outcome variables or where there was greater knowledge of the relation between outcomes and outputs (as in education), there was a greater weight attached to the outcome indicators in the performance agreements (70% of the overall weight). The flexibility of assigning different weights to outcome and output indicators made it easier for some sectors to enter into performance agreements, even if the sector didn't really know what their production process generating outcomes was and/or if the quality of the existing indicators was not fully satisfactory. The government then pre-announced that they would gradually increase the weights on the outcome indicators over time, thereby creating incentives for the sectors to learn about the production process.

The reward structure required that some minimum threshold (70% of the targets) be met before a sector receives any reward. Thereafter, the amount of the reward increased (up to a maximum) depending on the extent to which the threshold was exceeded. This type of reward structure was appropriate given that, at the time of entering into the agreements, there was incomplete knowledge on the part of both the principal and the agent of the costs and feasibility of meeting the targets. Allowing the agent some discretion as to which targets they can meet was perceived as fair and helps the buy-in for the arrangement.

The reward itself consisted of an additional month’s salary (as a maximum) or some fraction of an additional month’s salary for all members of the work unit. The State for Results was aware of the free rider problem associated with giving out group rewards. It managed that through the second stage of the Results Agreements, whereby each sector creates additional Results Agreements on specific results and with smaller teams. In smaller teams, it is more likely that moral suasion can be used to put pressure on potential shirkers. The State for Results does not attempt to measure the marginal contribution of a particular worker to the output and to reward that marginal contribution. The information costs to measure the marginal contribution would be very substantial and the effort could create disincentives for teamwork if workers are trying to maximize the measurement of their individual contribution.

The Ministry of Finance in Minas Gerais included a provision in the budget for financing the rewards because introducing Results Agreements generates a contingent liability. The liability was contingent in that the rewards may or may not have to be paid out. It is usually easier to allow for a contingent liability if the state or government is running a surplus. In that case, the cost is to accept a smaller surplus than would otherwise arise. This may be a small price to pay for generating more effective public sector. When a government is running a deficit, the cost of adding to the deficit can be nonlinear. And it is harder to convince the public that they are benefitting from a more effective public sector – as some may interpret (rightly or wrongly) the running of deficits as evidence of poor management.

3. Minas Gerais introduced what they called "**structural programs**" to bring new resources to help achieve targets.

The structural programs complemented existing programs and were designed to deliver the missing elements to the inputs needed to achieve results in the 11 results areas. Introducing new programs served, in essence, as a way to ensure that the increase in tax revenues generated from the crisis was not totally absorbed by increases in labor costs. The increase in public sector salaries had declined during the crisis period in the state and, if there had not been special Structural programs developed to help generate the results, it is quite possible that the pressure for salaries to recover could have absorbed much of the additional tax revenue generated from the recovery.

4. There was **strong follow-up on implementation.**

This was critical. There were monthly meetings on execution, annual reviews of progress in meeting targets and adjustment of targets. There was an annual publication of results. Indeed, there are two main ways that the introduction of Performance Contracts may lead to improved results. The first (and most obvious one) is that the Performance Contracts create a monetary incentive to achieve results. The second (and less obvious one) is that because there must be a review of the conditions to see whether results have been met, there is an opportunity for learning around the review process. Whether the learning actually takes place, of course, depends upon how the principal chooses to manage the review process.

5. Finally, the state reaped the benefits from having a cadre of **professionally trained public sector employees** distributed across different sectors (trained at Fundação João Pinheiro);

This made it easier for the management innovations to take hold within the sectors. In South Asia, there is also a tradition of developing trained public sector employees. Any attempt to introduce the type of management shocks in South Asia that were introduced in Minas Gerais would undoubtedly require rolling out result-oriented training in the public sector training institutes to accompany the reforms in the sectors and in the revamping of the relations between the Ministry of Planning/Finance and the sectors.

The results of the "Management Shock" have been impressive. For example, Minas Gerais had the highest increase of all states in Brazil in the percentage of 3rd grade students reading at the recommended level, as measured in national standardized tests. The success of the State for Results also contributed to the re-election of the Governor's party in the 2010 state elections. Following the elections, the state ended its "State for Results" program and, with the new administration launched its "Agenda for Improvement: Roads to Innovation in Public Sector Management". Many of the elements of the previous system have carried over, but there is now a greater focus on feedback to improve (to the point where they identify a task as Monitoring and Support for Execution rather than Monitoring and Evaluation).

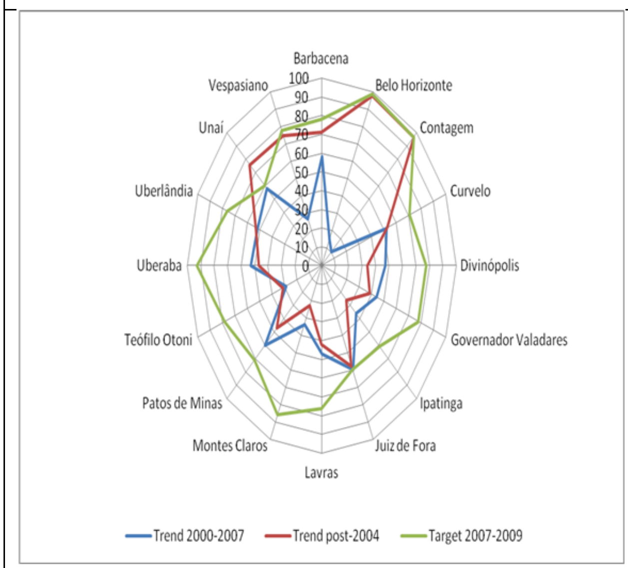
My final points have to do not with the financial instruments or the experiences of a particular state with a results approach, but on two tools that would likely be helpful in supporting a movement towards making the World Bank's South Asia portfolio predominantly results-focused. The first tool is the use of benchmarking to help judge the reasonableness of targets. In Minas Gerais (as would be expected to be the case in South Asian states and provinces) target-setting would be expected to be an integral part of generating a greater focus on results. As mentioned previously, targets are usually set through a process of negotiation between a principal (typically a Ministry of Finance of Planning) and an agent (typically a sector ministry). The outcome of this negotiation is critical to the effectiveness of the performance-informed budgetary process. If the negotiation results in an unreasonably high target which is subsequently not met, the principal has only two choices – either the requirement that the agent meet the target could be waived (in which case it damages credibility in the system) or the agent could be held to its agreement to meeting the target (in which case, being held to meeting an unreasonable target could breed resentment). There are, perhaps, less threats to the operation of the performance-informed budgeting system from setting a target that is unreasonably low, one that could be easily reached by the agents without doing anything different. While the problems of setting an unreasonably low target are not as visible as those associated with setting a target that is unreasonably high, there is still a cost. If nothing different is done – if there is no improvement in performance - then society is actually worse off because it takes real resources to mount the superstructure to support the performance-informed budgeting process.

In the benchmarking exercises that were carried out in Minas Gerais, the World Bank worked with the State for Results team and a group from Fundação João Pinheiro to analyze historical data on changes over time for indicators for which targets had been set. The data that are available in Brazil are extensive and we were able to look at performance data for different districts within the state over time. As a first step, we compared both the target and actual performance of a particular district to the distribution of all observed changes in the indicator of interest over all regions and time periods. Given the afore-mentioned costs, a target should not be set beyond what has historically been the best performance by any district over all time periods. However, it should certainly be set to generate a stretch target. In the second step, we attempted to respond to a possible reaction made by a district to being compared to all other districts. They could object to being compared to districts that were larger (or smaller), wealthier (or poorer). To respond to that concern, we took note of the characteristics of the district by comparing both the target and actual performance to a counterfactual conditional distribution. The counterfactual conditional distribution was obtained by estimating 99 separate quantile regressions relating the change in the indicator of interest to a set of observable variables. This uses data on all districts over all time periods. Then one multiplies the actual values of the observable variables for the district by each of the 99 sets of coefficients to yield a counterfactual distribution of predicted values. Then one can compare both the target and the actual values to this counterfactual conditional distribution.

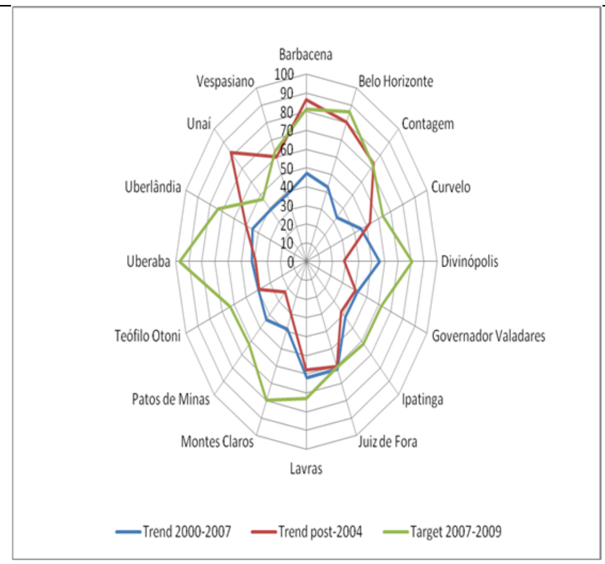
Figure 3: Compares the actual performance and the target in homicide reduction in different police districts in Minas Gerais to the empirical distribution and the counterfactual conditional distribution.

Figure 3.

Unconditional relative performance on homicide reduction between 2000-2007, by Police Regions



Conditional relative performance on homicide reduction between 2000-2007, by Police Regions

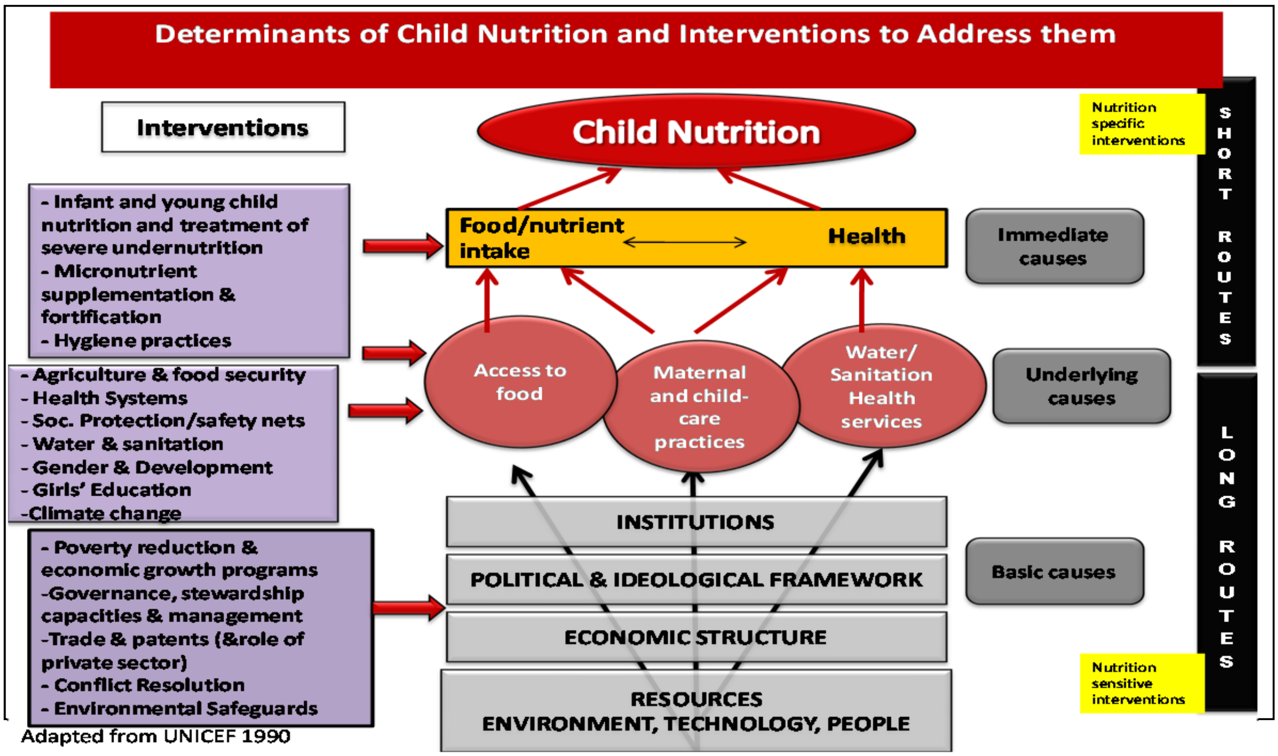


It is worth noting that the differences between the unconditional and the counterfactual conditional distributions are not all that great. It would, perhaps, be disturbing if they were. One might be inclined to conclude that this ability to take note of a country's initial conditions does not add much. However, the value of being able to calculate the conditional distributions has more of a tactical then a technical value. The ability to take note of initial conditions makes it harder for a country (or state or district, depending on the unit of analysis) to dismiss the unconditional information as being completely irrelevant for them because of some factor which was not taken into account. The value is the ability to put on the table what has been achieved and to ask why it would not be possible for a country to achieve what others have achieved.

The development of the second tool responds to Ravi's call for greater clarity in project documents on outcomes, and on evidence for the causal link from project inputs to outcomes. A team from the World Bank and the REACH Initiative are working on developing a Multisectoral Simulation Tool for nutrition under the South Asia Food and Nutrition Security Initiative to help increase operational understanding of what might be driving nutritional outcomes in Bangladesh and India and what might be required to scale up nutrition interventions to reach desired targets. This is being done to support the SUN (for Scaling Up Nutrition) Initiative, a larger worldwide initiative.

In nutrition, the causal chain involves multiple sectors and can be complex. It is important to try to manage that complexity. Fortunately, there are tools available that have been developed in other fields (notably System Dynamics[6]) that show promise for building operational understanding of the causal chain. This approach is briefly described below. The starting point is the UNICEF framework, which was introduced in 1990 and has long emphasized the importance of taking a multisectoral approach. A recent representation of the UNICEF framework is given in Figure 4.

Figure 4: A Recent Representation of the UNICEF Framework for Nutrition



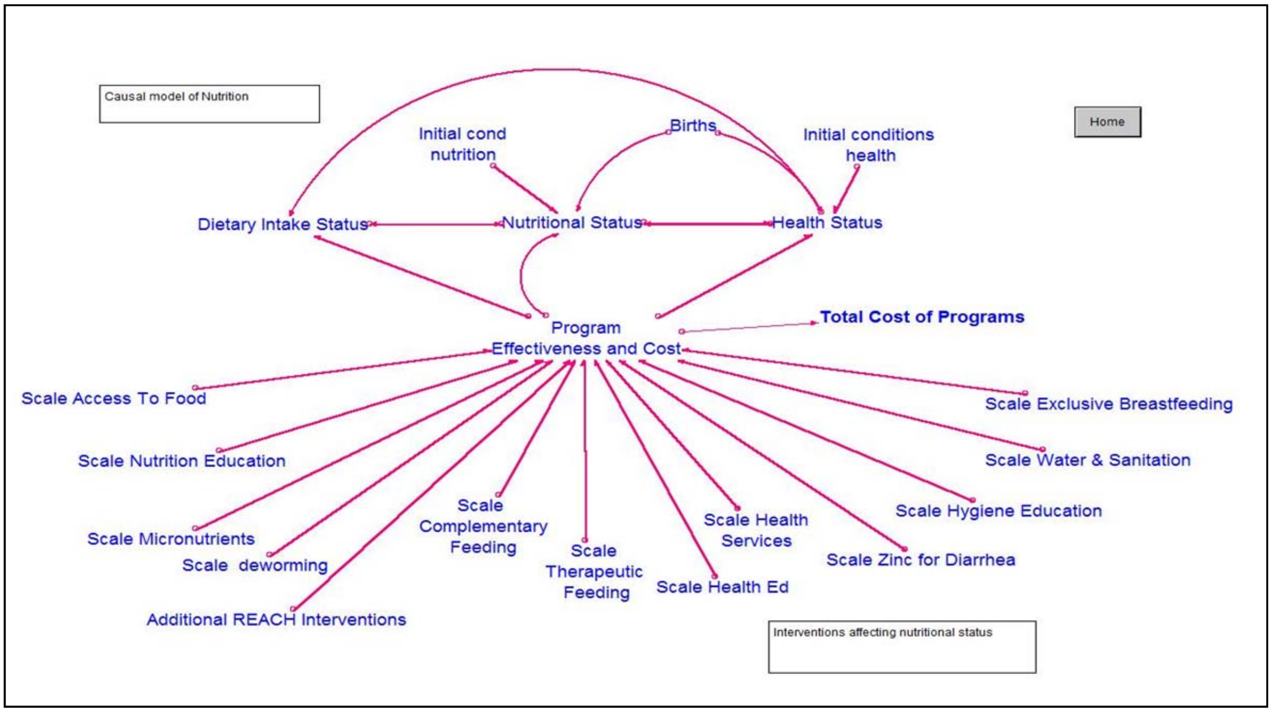
Source: The Neglected Crisis of Under nutrition: DFID's Strategy, March 2010

This framework clearly illustrates the multisectoral nature of nutrition, but does not really help countries come up with quantitative estimates of what a district or country may need to do. For that, one needs a more operational version of the UNICEF framework that makes explicit links between some of the activities and the outcomes and takes account explicitly of a district or a country's initial conditions.

6 For an interesting introduction to System Dynamics see, Richmond, J. et al (2010), "Tracing Connections: Voices of Systems Thinkers", I see Systems and the Creative Learning Exchange.

Figure 5 illustrates an initial high-level map from the Multisectoral Simulation Tool[7]. A high-level map is a device that allows one to see the forest. This particular high-level map shows that there is a relationship between health status and nutritional status and that the effect goes in both directions. It also shows that this interaction depends upon the outcome of births and on the initial conditions associated with the drivers of nutritional and health status. Finally, the high-level map identifies some specific interventions which, depending on their program effectiveness and scale, would be expected to affect nutritional status directly or indirectly acting through health status[8].

Figure 5: High Level Map from Multisectoral Simluation Tool for Scaling Up Nutrition



From this high-level map, the user can drill down to look at the trees – the detail needed to understand the causal model determining nutritional outcomes or simply the nuts and bolts of what resources may be required to increase the scale of a particular intervention. This is the way one breaks a complex problem down into more manageable chunks. However, what goes on in one box affects what happens in another and the feedback can go in both directions. For example, those children who are malnourished are more vulnerable to illness, but those children who are sick are also more likely to lose nutrients and go from being adequately nourished to being malnourished. This is a negative feedback loop that is difficult to handle in many models, but not in System Dynamics models. Indeed, System Dynamics was developed precisely to deal with feedback loops.

[7] The MST is developed using I-Think software, a specialized package for System dynamics modeling (see www.iseesystems.com).

[8] The interventions included in this high level map are the ones featured in Horton, S. et al, "Scaling Up Nutrition: What Will It Cost?", World Bank, 2010

The individual interventions are taken from the SUN list of interventions. Within each box, one would have to go into some detail describing what the initial conditions are and what resources would be necessary to scale up. The box on effectiveness is a placeholder where one would define explicitly what is known about the effectiveness of the different interventions. If the effectiveness is not known or is known only with a high degree of uncertainty, one can work with ranges of estimates and see how, in the simulation, the outcomes vary according to the estimates of the effectiveness.

It is important to recognize that the work will involve an iterative process. After some initial work making explicit the mental models that different actors have and trying to come up with a shared vision of what is driving the system, the interventions will take place and the initial results will become known. If the results are not what was expected, the first thing to check is whether the planned actions were actually carried out. If they were carried out but results were still unexpected, it could be that the parameter values were not what they were believed to be. If the interventions were carried out as planned and the parameter values were largely as expected, it is possible that the system driving the results does not work as thought. In that case, one would revisit the explicit models of how the system is characterized to work.

The hypothesis behind this approach is that a more systematic approach[9] may allow a country or district to reach a goal in 5 years instead of 10. In the case of malnutrition in South Asia, this would translate into millions more children growing up with an enhanced ability to learn and a reduced risk of dying.

Conclusions

The challenge that Ravi Kanbur posed, for the South Asia Region to have a large results-based financing portfolio in the coming few years, is ambitious. If this challenge had been made even five years ago, it would be unrealistic to think it could be met. But there have been recent advancements in lending instruments. There have been valuable lessons learned on how to implement practical and flexible Performance Contracts through the experience with places such as Minas Gerais. There are new tools coming on line now or in the near future. To be sure there are some considerable obstacles to overcome to advance on the results agenda in South Asia (for example, the data systems are not as well developed as in Brazil). However, there is certainly the potential to meet that challenge and, in so doing, to help South Asian countries make faster progress in achieving their goals.

[9] This approach follows essentially the Plan, Do, Study, Act cycle advocated by Deming (following Shewart, see http://en.wikipedia.org/wiki/PDCA). It also follows the logic of the World Bank's own project cycle. The difference is that this approach attempts to make more explicit the mental model of what is often only held implicitly.

OTHER A.E.M. WORKING PAPERS

WP No	Title	Fee (if applicable)	Author(s)
2012-02	Aid to the Poor in Middle income Countries and the Future of IDA		Kanbur, R.
2012-01	Does Kuznets Still Matter?		Kanbur, R.
2011-21	Meeting multiple policy objectives under GHG emissions reduction targets		Boisvert, R., and D. Blandford
2011-20	The Theory of Biofuel Policy and Food Grain Prices		Drabik, D.
2011-19	Factors Influencing Adoption of Integrated Pest Management in Northeast Greenhouse and Nursery Production		Gómez, M.
2011-18	Urban Agglomeration Economies in the U.S. Greenhouse and Nursery Production		Gómez, M.
2011-17	Evaluating the Impact of the Fruit and Vegetable Dispute Resolution Corporation on Regional Fresh Produce Trade among NAFTA Countries		Gómez, M., Rizwan, M. and N. Chau
2011-16	Does the Name Matter? Developing Brands for Patented Fruit Varieties		Rickard, B., Schmit, T., Gómez, M. and H. Lu
2011-15	Impacts of the End of the Coffee Export Quota System on International-to-Retail Price Transmission		Lee, J. and M. Gómez
2011-14	Economic Impact of Grapevine Leafroll Disease on Vitis vinifera cv. Cabernet franc in Finger Lakes Vineyards of New York		Atallah, S., Gomez, M., Fuchs, M. and T. Martinson
2011-13	Organization, Poverty and Women: Andhra Pradesh in Global Perspective		Dev., S., Kanbur, R. and G. Alivelu
2011-12	How have agricultural policies influenced caloric consumption in the United States?		Rickard, B., Okrent, A. and J. Alston
2011-11	Revealing an Equitable Income Allocation among Dairy Farm Partnerships		Dressler, J. and L. Tauer

Paper copies are being replaced by electronic Portable Document Files (PDFs). To request PDFs of AEM publications, write to (be sure to include your e-mail address): Publications, Department of Applied Economics and Management, Warren Hall, Cornell University, Ithaca, NY 14853-7801. If a fee is indicated, please include a check or money order made payable to Cornell University for the amount of your purchase. Visit our Web site (*http://aem.cornell.edu/research/wp.htm)* for a more complete list of recent bulletins.