# The Rise of Measurement-Driven Governance:
# The Case of International Development

Jacqueline Best
University of Ottawa

Note: this is a pre-publication version of an article forthcoming in
*Global Governance*, 2017.

## Abstract

Global governance is increasingly about measuring, ranking and scoring. Focusing on the case of international development, this paper suggests that we can learn a great deal about the recent changes in global dynamics by examining the rise of measurement-driven governance. This article engages with the recent scholarship on "new governance," particularly the emerging literature on more experimentalist forms of governance. This study finds some evidence of a more experimentalist kind of governance in international development, but suggests that the specific technologies of measurement and accountability through which these new structures of decision-making are being facilitated are far from politically neutral. They are instead sources of considerable power that, in many instances, reinforce existing asymmetries and undermine the deliberative potential of these governance practices.

Keywords: measurement, experimentalist governance, international development

## Introduction

Global governance is increasingly about measuring, ranking and scoring. In contexts as varied as international development, anti-money laundering regulations, global environmental governance and international finance, we see a growing emphasis on rating good and bad performers, developing scorecards and black lists, and then publishing the results in the name of transparency and accountability. While a growing number of scholars have begun to analyze the complex forms that global governance has begun to take in recent decades, they have not paid enough attention to their growing reliance on techniques for measuring performance and results.

Focusing on the case of international development, this article brings the emerging scholarship on the politics of measurement[1] into conversation with the literature on "new governance,"[2] suggesting that we can learn a great deal about the recent changes in global governance by examining the specific but widespread measurement techniques that define its everyday activities. I will focus in particular on the literature on more "experimentalist" forms of governance, whose advocates argue that we are witnessing the spread of a more decentralized, deliberative and recursive form of governance in which local actors gain greater decision-making authority in return for various forms of monitoring and peer review.[3] This article finds some evidence of a more experimentalist kind of governance in international development, but suggests that these more deliberative forms of decision-making have gradually become distorted through their reliance on measurement techniques that produce perverse policy outcomes, reinforce existing power asymmetries and undermine deliberative dynamics.

Much of the literature on new forms of global governance has focused on either the EU or on major international issue areas whose management is dominated by industrialized states. This article in contrast looks at the other end of the global spectrum of power and privilege, examining the changing practices of international development. The governance of international development has changed in important ways since the days of the "Washington Consensus." After a series of failures in the 1990s, development actors began to redefine their policies around the pursuit of "aid effectiveness." Among the new policies that development organizations adopted in this context was a series of measures designed to foster greater country ownership and a widespread embrace of results management—policies that demonstrate, in different ways, the key characteristics of both experimentalist and measurement-driven governance.

Focusing on the case of international development is useful for several reasons. Results-based management has a particularly long history in this field, providing useful insights into its likely effects in other governance contexts. At the same time, as one would expect, the power dynamics involved in recent governance changes are particularly stark—making them more obvious (but not necessarily more significant) than in other cases. This article's conclusions therefore speak primarily to the changing character of international development practices. Yet the problematic consequences of their increasing reliance on measurement-driven governance strategies also has lessons for other issue areas, where we also see a trend towards measurement-driven governance.

I begin this article by engaging more thoroughly with the literature on "new governance," focusing in particular on the recent work on experimentalist governance. I then trace the evolution of development governance from the late 1970s to the mid-2000s, as aid organizations shifted from structural adjustment to the pursuit of country ownership and aid effectiveness. I suggest that a number of the policies associated with this shift contained key elements of experimentalist governance yet did not achieve their deliberative potential. In order to understand these shortcomings I then look at the recent spread of results measurement techniques and consider their impact on efforts to foster country ownership. I conclude by considering these techniques' broader implications for efforts to foster more deliberative forms of governance: the growing emphasis on measurement-driven governance tends to narrow the forms of deliberation and accountability to what can be counted, impoverishing rather than nourishing genuine political debate.

## From experimentalist to measurement-driven governance

The landscape of global governance has changed. The traditional international organizations are still around, but the terrain has become crowded with new actors, arrangements and techniques. A growing literature has emerged in response, as scholars set out to describe these new patterns of governance: as multi-level governance, networked governance, regime complexes and experimentalist governance, to list a few.[4] As their names suggest, most of these literatures seek to capture the *shape* of these emerging patterns of governance, arguing that the days when global governance was dominated by states and major international organizations have passed: much of the real work of governance is done through multiple, overlapping groups of actors and regulations.

While there has thus been considerable effort in the new governance literature to trace the *form* of these new kinds of governance, there has been less attention paid to *how* they work—particularly at the mundane level of everyday practices.[5] One exception to this inattention to how governance works is the literature on experimentalist governance. In spite of its use of the spatial vocabulary of "architecture," this scholarship focuses on changes in *how* decisions are made and knowledge is produced, suggesting that we are witnessing a shift to more open-ended and recursive form of governance.[6]

Charles Sabel and Jonathan Zeitlin have proposed a four-part model of how experimentalist governance works: a framework with open-ended goals is agreed by all parties; the implementation of that framework is delegated to lower-level actors who have significant discretion; those local actors provide feedback on how things are working out on the ground; and the framework itself and key practices are reviewed and revised on a regular basis.[7] Experimentalist governance, they suggest, is not only more functional in an increasingly complex and uncertain policy environment, but is also potentially more democratic. Even if the local actors who have been delegated more authority are not directly elected, they are accountable in different ways: through ties of horizontal accountability that are made possible through peer review.[8]

In this article, I will build on many of these insights from experimentalist governance into how these new forms of governance work, while also playing more attention to the specific practices through which the processes of accountability and evaluation operate. In doing so, I will focus on four key features of these emerging practices of governance: the *delegation of authority* to local actors, the development of *mechanisms of accountability* for those actors, the application of new *techniques of measurement* in order to evaluate policy outcomes, and the *transformation of power dynamics* as a result. Together, these emergent practices constitute an increasingly measurement-driven style of global governance.

As it has become apparent that problems such as global poverty, climate change and financial instability resist straightforward solutions, one of the ways that international institutions have responded is by *delegating authority* to local actors—whether they are expert committees, as in the European Union, or borrower governments, as in the case of international development.[9] The reasons for this transfer of authority are both pragmatic and political. It is assumed that because these actors are closer to the ground, they will be more sensitive to local contexts and better able to respond to the complexities of specific situations.[10] At the same time, the transfer of authority is also often an effort to strengthen institutional legitimacy, whether by reinforcing the deliberative dimensions of decision-making in the EU,[11] or by fostering greater ownership by domestic political leaders in development programs, as I will discuss below.

In outlining the specific characteristics of experimentalist governance, its advocates have tended to use principal-agent theory as a foil. From the perspective of principal-agent theory, the move to delegate authority would be seen as a risky gambit since it reduces the control exercised by the principals—whether EU institutions or development organizations—over the agents to whom they have granted this authority. Yet, the experimentalist governance literature suggests that the usual principal-agent dynamics do not apply here: because the issues being addressed are so complex and the results so uncertain, neither principals nor agents actually know their preferred outcome until they have tried a few different options: delegation and experimentation can help principals identify and achieve their preferences.[12]

There do nonetheless remain risks for the principals. One significant risk is that accountability will be weakened because of the distance between central organizations and local actors, and because of the authority that has been delegated to them (often without the stamp of electoral approval). In response, governance actors have developed a set of *mechanisms of accountability*: based on transparency, publicity and peer review, they promise to open the actions of local actors to the scrutiny not just of the principals, but also of other local actors. These mechanisms often bring civil society actors into the process—generating knowledge, providing feedback and fostering a more democratic form of governance.[13]

There remains a further risk for international organizations, however—one that is not explicitly addressed in the experimentalist literature. We might think of this as a methodological, even epistemological, dilemma: this is the risk that the complexity and uncertainty of the policy environment will make it difficult to know what the actors on the ground have actually done, and what the effects of their actions have been—knowledge that is necessary to hold local actors accountable.[14] In response, international organizations have developed a whole range of *techniques of measurement*, including various forms of results and performance evaluation: all designed to produce more and better knowledge about governance processes and outcomes.

So far, this overview of measurement-driven governance is a straightforward, even optimistic, liberal story about the transformation of global governance through delegation, better knowledge and accountability—a story that is not so different from the other accounts of networked, multi-level, and experimentalist governance. Yet if we look more closely, we can also identify particular *power dynamics* at stake in these new practices of delegation, accountability and—above all—measurement. Sabel and Zeitlin do point to the central role of performance indicators in the EU, and also discuss their centrality in recent American educational reforms, where standardized testing and the ranking of schools is essential to their management.[15] They do not, however, examine the political dynamics of such measurement techniques.

In experimentalist governance situations, hierarchical principal-agent dynamics are supposed to be replaced by polyarchic ones, in which principals can no longer impose their preferred solutions, and peers discipline each other through horizontal forms of accountability. The main hierarchical source of power, if it can be called that, is the "penalty default"—the threat that, if actors do not manage to discipline each other effectively, they will be forced to accept an outcome that they all hope to avoid; in formulating such a penalty, it is assumed that the "rule-maker" will pursue the general good by formulating incentives that will encourage actors to produce the optimal outcome.[16]

This article argues, in contrast, that the techniques of measurement put into place in many experimentalist governance arrangements are a source of power that can provide advantages to certain actors. The fact that principals cannot impose their will unilaterally does not mean that they cannot exercise a considerable amount of influence over the process and outcome. At the same time, the experimentalist governance literature is right to point to the serious limitations of straightforward principal-agent frameworks in such contexts. Yet this does not mean that power is less present than in the past—just that it has taken different forms.[17] As I will elaborate in my discussion of development governance, although benchmarks, performance indicators and results matrixes may appear technical and politically neutral, they are crucial tools for gaining knowledge about—and thus control of—governance processes.

Measurement-driven and experimentalist forms of governance are therefore best understood as overlapping sets of practices, like a Venn diagram. It is not difficult to think of examples of measurement-driven governance that do not seek to foster the deliberative dynamics that are central to experimentalist governance (many rankings and blacklists for example involve no consultation with those affected). There also exist forms of experimentalist governance that do not rely on results and performance-based metrics (as I will discuss below, early participatory forms of development aid involved much more qualitative and open-ended forms of evaluation and feedback). Yet, as such measurement techniques become increasingly widespread, the overlap between the two categories is also growing, as various forms of experimentalist governance become subject to increasingly quantitative systems of measurement and evaluation.

## From structural adjustment to country ownership

Today's increasingly delegated and measurement-driven approach to development governance contrasts with the structural adjustment era of the 1980s and 1990s. First introduced at the very end of Robert McNamara's tenure as World Bank President and championed by his successors, structural adjustment lending provided long-term funds in return for requiring specific "structural" policy changes by borrowing countries—changes that agency staff believed were essential for getting those countries back on solid economic ground. This was a relatively direct approach to influencing countries' policies—and thus of controlling (or at least trying to control) economic outcomes. Yet this direct style of governance was not as successful as its advocates had anticipated. By the late-1990s, there were increasing signs of the policies' failure.[18] In 1999, the UNDP declared the 1980s a "lost decade" in Sub-Saharan Africa, in spite (or because) of the efforts of development organizations.[19] At the World Bank, the organization's own internal evaluations were indicating a significant decline in success rates.[20] And at the IMF, the staff was facing considerable resistance to its policies in the aftermath of the

Asian financial crisis—a crisis that they had failed to predict and were blamed (by World Bank leaders no less) for having made worse.[21]

Aid agencies thus faced both technical and political crises in the late 1990s: after their programs had failed so publicly, they could no longer pretend that they had a monopoly on development expertise. At the same time, borrowing country governments and non-governmental organizations (NGOs) were becoming increasingly vocal in challenging the popular legitimacy of international development agencies, arguing for a more democratic, participatory approach to development assistance. This double challenge to the institutions' expert and popular authority produced a broad-based debate about policy failure and the meaning of development success, ultimately leading to the redefinition of development practices around the idea of "aid effectiveness."[22]

Although the term was popularized by the OECD's 2005 Paris Declaration on Aid Effectiveness,[23] this approach to development emerged in the mid-1990s among policy-oriented academics and at the World Bank, and soon became a touchstone in many different development organizations, think tanks and NGOs. One of the major differences between the structural adjustment approach and the aid effectiveness agenda was the increasing recognition that development problems are highly complex and, more importantly, that they are not just economic, but rather require the right social and political institutions—and a sufficient degree of local political "ownership"—to be addressed successfully.[24] Without this ownership, it was discovered, no number of conditions—the principal means through which donors had been exercising their influence over borrowers—is likely to produce the desired outcome.[25]

Ultimately, the idea of country ownership not only posed a challenge to aid agencies, but also provided a potential solution to their declining legitimacy, as organizations sought to get borrowing governments and civil society groups more involved in the processes of defining the problems and implementing the solutions. This strategy of country ownership manifested itself in a range of different specific policies. The World Bank and IMF replaced their Policy Framework Papers, which had been developed exclusively by their own staff, with Poverty Reduction Strategy Papers (PRSPs), which were to be developed by the borrowing country in consultation with civil society. The IMF, the World Bank, and several donors including the UK and many Nordic countries also "streamlined" their conditionality in an effort to reverse the proliferation of conditions that had occurred during the structural adjustment era.[26]

Many donors also adopted general budget support as one of the main forms of aid delivery, beginning in the late 1990s in conjunction with the World Bank and IMF's PRSPs. As its name suggests, this kind of support provides financial assistance directly to a country's general budget, rather than earmarking funds for specific purposes. Because the aid is transferred directly to a borrowing country's budget, development agencies must rely on that country's own financial management systems to ensure that the money is well-spent.[27] At the same time, there are various agreements and forms of monitoring in place to allow donors some control over the process.[28]

The redesign of international aid policies around the concept of ownership that occurred in the late 1990s looks a lot like experimentalist governance. As aid organizations recognized that the problems of development were complex and that their solution depended on borrowing governments' support, the power dynamics underpinning development finance shifted: to borrow the language of the experimentalist literature, once they recognized that borrowers' ownership was vital to success, aid agencies no longer had "the capacity to impose [their] own preferred solution without taking into account the views of the others."[29] Such solutions were also subject to greater uncertainty given the increasing complexity of aid challenges. In this context, the PRSPs serve as a mutually agreed framework between the donor and borrower, focused on the common goal of poverty reduction. When funds are allocated through general budget support, considerable decision-making authority is delegated to local actors. Annual progress reports, derived from the various forms of measurement and evaluation built into aid programs, are designed to allow for feedback from local actors—particularly civil society groups who, it is hoped, will use the published accounts of their government's progress towards key indicators as a way to hold them accountable. There is also a mechanism for revising the basic framework, since the PRSPs are agreed for a set period of time and have to be revised every three to five years.[30]

Gráinne de Búrca and other experimentalist governance scholars have pointed to the PRSP as a potential model for a more democratic, recursive form of governance.[31] At the same time, they have also noted some of the serious shortcomings of development organizations' efforts to foster country ownership. Both internal and external reports have raised questions about the degree to which borrowing governments have genuinely taken the initiative in developing the poverty reduction frameworks as well as the extent to which civil society groups have played a meaningful role in setting development priorities.[32] In spite of the early enthusiasm about the various policies designed to foster greater country ownership, over time, staff working within aid agencies, developing country governments, NGOs and scholars have become increasingly skeptical about their benefits.

Why was this effort to generate a more deliberative and recursive form of development governance not entirely successful? De Búrca focuses on the failure of the IMF and World Bank to leave the initial framework sufficiently open to revision and their refusal to import some of the same commitment to openness and deliberation into their own governance processes.[33] Yet there is a further factor that played an important role in undermining these deliberative processes: the central and growing reliance of aid agencies on certain techniques of measurement, which ultimately led them to resist some of the more open-ended and deliberative aspects of these policies and to reassert a more effective form of control over development processes.

## The turn to results measurement

Although development agencies were initially quite enthusiastic about efforts to build country ownership and engage in bottom-up institutional reform, they soon encountered a dilemma: agency staff knew that to be effective, aid interventions needed to be longer-term and focused on subtler and more complex process such as institution-building rather than on more visible projects like building schools or roads; yet, agency staff also needed to be able to demonstrate the effectiveness of their efforts to often-skeptical politicians and voters. The solution that they ultimately turned to in order to resolve this dilemma was to find ways of demonstrating the *results* of their own and borrowing governments' actions: they hoped to draw a causal line between agencies' programs and specific improvements on the ground.[34]

Results management techniques have a long and contested history in development circles. Although they came to prominence through the global spread of new public management practices in countries like New Zealand in the 1980s and 1990s, they emerged over a decade earlier in aid agencies. The United States Agency for International Development (USAID) first adopted the logical framework or "LOGFRAME" approach to project management in 1969.[35] This LOGFRAME was one of the first results matrixes, a framework that asked project developers to identify not just the inputs into development programs but also the anticipated outputs and the indicators they would use to measure their achievement. This was a new and powerful technology of measurement and evaluation that sought to draw a causal line between policymakers' actions and their consequences.

By the early 1990s, this interest in policy results took on a more pronounced political flavour. New public management advocates, inspired by public choice theory and new institutional economics, sought to transform the public sector so that it worked more like the private sector.[36] Believing that individuals are always motived by rational self-interest, these management experts were skeptical of the traditional emphasis on the public good as the motivation for government bureaucrats. They introduced a series of contracts among government departments and individual staff in order to individualize responsibility and to make each accountable for the results of their own actions. Results measurement was thus both a powerful mechanisms of accountability and a new means of control. Replacing social ties of responsibility with contracts linked to results delegated some of the initiative to public sector staff while also promising government leaders the information they needed to monitor and manage their actions.

Although development agencies had been in the forefront of very early results-based initiatives in the 1960s, by the mid-1990s and early 2000s, many staff were aware of the limits of translating complex development processes into tidy causal chains linking inputs, outputs and results. Many agencies sought to develop more experimental and less measurement-driven approaches to program development and assessment. A number of agencies including the World Bank sought to build on the qualitative "Participatory Rural Appraisal" techniques popularized by Robert Chambers.[37] Staff at the Canadian International Development Agency (CIDA) outlined the risks involved in a simplistic approach to results and advocated a recipient rather than donor-driven strategy.[38] The British Department for International Development (DFID), for its part, sought to find creative ways of evaluating more complex and long-term development priorities by relying more on qualitative assessments in tracking its results.[39] This more qualitative approach to results-measurement did not last long however, as politicians wanted to be able to demonstrate clear, measurable results to their constituents.[40] By the time the Paris Declaration was agreed in 2005, quantifiable results had been identified as one of the five core pillars of aid effectiveness.

### Recent results measurement initiatives

By the mid-2000s, results measurement had come to be seen as a crucial complement to the emphasis on country ownership—a means to ensure that donors' expenditures are effective and borrowers accountable. If we are to grasp the influence of this new measurement technology in development policy, we only have to look at the World Bank, which has rapidly adopted results-oriented techniques in its own programs and has also promoted the results agenda more widely among its members. Its International Development Agency (IDA), which provides subsidized loans to low income countries, has had a strong results orientation since 2002.[41] Beginning in 2011, Bank began to integrate some of the same metrics into its lending to middle-income countries through its Corporate Scorecard and the annual *World Bank for Results* publications, which provide an overview of country-level results related to Bank projects.[42] The Bank also now includes results frameworks in all of its Country Assistance Strategies, lending projects and programs, and, in 2012, adopted an entirely new financing instrument, the Program-for-Results, which is more closely linked to results monitoring and measurement.[43]

While the Bank's turn to results has been analyzed by scholars,[44] less attention has been paid to the wide range of other institutions that have followed the same path. In recent years, the Millennium Challenge Corporation, the European Commission, and the Center for Global Development have all played important roles in the turn towards results. These three examples include some of the most influential results-based initiatives and demonstrate the range of different kinds of organizations and policies involved.

The Millennium Challenge Corporation (MCC) was the first agency wholly structured around the principles of selectivity and results measurement and, as such, became the "elephant in the room" that the other development organizations felt they had to emulate.[45] First created by the US government in 2004, the MCC has a relatively narrow mandate: through five-year contracts, or "Compacts," that seek to reduce poverty by increasing incomes, the MCC sets out to help countries that have already demonstrated the capacity to use aid effectively. A country's eligibility for assistance depends on its prequalifying according to a set of quantitative indicators of that country's economic, democratic and institutional credentials.[46] At the same time, the MCC delegates a certain amount of decision-making authority to government actors through its commitment to country ownership.[47]

In an effort to develop adequate knowledge about those results, the MCC staff have placed considerable emphasis on new techniques of measurement. Not only are decisions about eligibility made through various indexes of countries' progress, but the outcomes of individual MCC investments are also tracked. In this respect, the MCC is quite rigorous in its efforts to determine the effects of its policy interventions, making use of control groups and counterfactual analysis to try to determine what were the causes of the ultimate outcomes.[48] The MCC has also developed several accountability mechanisms, publishing their actions and outcomes at various steps along the way and encouraging civil society groups to use that information to hold both governments and the MCC itself accountable.[49]

The European Commission (EC) also pioneered a results-based approach in 2000, in a program that combined general budget support with various forms of monitoring and evaluation: the Good Governance and Development Contracts. As with the MCC, there is a basic framework that borrowing countries must agree to at the outset, and a country must also meet certain prior economic and institutional criteria and demonstrate its commitment to certain European Union values, including human rights, democracy, and the rule of law.[50] Within these parameters, considerable authority is delegated to borrowing countries to decide how to spend the funds. At the same time, although governments receive some of the aid up-front once they have met the prior conditions, the rest is delivered through performance-based tranches linked to a country's progress in obtaining key results.

One of the results-based approaches to aid delivery that the EC has begun to explore, particularly for countries not ready for a full "good governance contract," is Cash on Delivery. The idea of Cash on Delivery (COD) has been championed by Nancy Birdsall and Owen Barder at the US-based NGO, the Centre for Global Development, and piloted in a number of African countries. In an early paper on the topic, Birdsall and Barder argued for the need to develop a form of results-oriented aid that would provide incremental rewards for progress towards the agreed goal (in its initial incarnation, the approach was called "payment for progress").[51] COD aid takes the form of a contract between the donor and recipient, in which the donor commits to paying a certain amount for each "unit" of the desired result. For example, the UK's DFID has a pilot project in Ethiopia in which it has contracted to pay £50 - £100 for each student who sits and passes a particular exam, which it treats as a proxy for improved educational outcomes.[52]

In Birdsall and Barder's conception of COD assistance, the recipient government is granted full authority for deciding how to achieve the desired results: "The approach would give flexibility and autonomy to local institutions, providing an opening for local institutional experimentation, while at the same time ensuring that aid pays only for real, measurable achievements."[53] The measurement of those results becomes absolutely essential, of course, since the transfer of funds depends on that calculation. To ensure their validity, the types of results generally used as indicators are straightforward and easily quantifiable, and are verified by a third party. The simplicity of these indicators is also designed to make it easier for domestic civil society groups to hold governments accountable for their performance.[54]

## The politics of results measurement

What do these three examples of results-based initiatives, together with the mixed success of the PRSP, general budget support and other efforts to foster country ownership, tell us about the impact of new measurement techniques on the more experimentalist aspirations of development governance? They point to the trade-offs involved in delegating authority, the perverse impact of results measurement on policy, the persistence of power asymmetries, and the instrumentalization of accountability and deliberation.

All three of these results-based initiatives delegate some decision-making authority to local actors—in this case, recipient governments—whether through the ownership of program choices with the MCC, general budget support with the EC, or complete autonomy over policy decisions with COD. Yet there is a trade-off for this transfer in authority: in each case, the borrowing country must score high enough in the performance evaluation. In the case of the MCC, a great deal hinges on the initial evaluation of a country's score as a likely candidate for development success. For the EC program, a borrowing government has the flexibility offered by general budget support, but must demonstrate that they are on target to meet specific results in order to receive later disbursements of aid. With COD, every dollar of aid is linked to the calculation of the results achieved—individuals immunized or students passing an exam. Through their use of these forms of evaluation, development organizations have effectively shifted the form of their control, from a direct set of ex-ante conditions on borrowers' economic and political policies, to an indirect set of "process conditions" and ex-post evaluations of their success or failure.

This increased measurement and monitoring of countries' performance matters because efforts to measure results do not deliver the kind of accurate knowledge that they promise, and because they often distort the policies that they pretend to support. Results-management hinges on the belief that it is possible to establish a results-chain: a causal chain that links specific policy inputs (like an increase in nurse training) to both short-term outputs (more nursing care in the community) and longer-term outcomes (lower infant mortality rates). Yet, as many development organization staff will admit, attributing outcomes to specific policies is incredibly difficult.[55] Even the MCC, which has been innovative in this respect, has struggled to accurately measure results—particularly in a way that can be standardized across different programs and countries.[56]

Focusing on measurable results can distort policies so that they target only the kinds of outcomes (test-taking, immunizations, income increases) that can be easily counted. As one OECD staff member put it, "with an extreme focus on bean counting now, there might be a tendency to move away from the more difficult areas like institution building and governance."[57] Moreover, as one narrows the scope of policies to such measurable outcomes, it is much easier to produce perverse incentives—to encourage governments to shift resources into those areas where results are being measured and away from those other, equally important, areas where they are not.[58]

These techniques of measurement also contribute to power asymmetries between donor and recipient. In each case, the capacity to quantify, score and publish a recipient's performance is central to the donor's strategy for keeping them on track with desired policy reforms. Donors as principals may no longer be able to impose their preferences without some political support from recipients—but they nonetheless retain important forms of control over the development process. The benchmarks and results indicators that various groups use to hold local governance actors accountable are not neutral, but have their own often-powerful effects. For example, the "MCC Effect" leads needy countries to make policy changes before they even obtain aid, in the hopes of scoring better on the MCC indexes and qualifying for its scarce resources.[59] The World Bank's "Doing Business" indicators play a similar role, as countries discipline themselves in an effort to move up the rankings. Although the MCC and World Bank both present these indexes as neutral and universal, they of course reflect particular (western) values, and thus become an effective way of influencing poorer countries' policies.

This preoccupation with measurement also has consequences for the kind of country ownership, horizontal accountability and deliberation that are supposed to be enabled by these new forms of governance. Many of the governance initiatives that rely on these new forms of measurement are ultimately driven by donors' rather than recipients' concerns and needs, compromising any country ownership. MCC programs, for example, must be able to demonstrate measurable results within five years—a timeframe not designed to respond to recipients' needs but to meet its reporting requirements to Congress.[60] The recent ramping up of results-focus is primarily aimed at satisfying donor publics: as a recent EC report put it, "Improving the visibility of EU action and better communication of results will help maintain public support for development."[61] Although earlier experiments with results-based approaches sought to increase programs' responsiveness to local communities' needs, more recent results-oriented approaches to development have been aimed more at reducing the risk of electoral defeat in donor countries.

Agencies' preoccupation with measurable results has also made them more wary of messier forms of consultation and accountability. In the case of the PRSP, although development staff were initially enthusiastic about the policy, they quickly discovered that there was a tension between their desire for more participatory processes and their need to show measurable results. As one former senior World Bank staff member described PRSP consultations:

> If the process was locally-driven, it often ended up being mired in domestic politics, and most governments took the easy way out which was letting everyone throw in the kitchen sink: there was no prioritizing or costing, and it risked becoming a useless exercise.[62]

These technical failings of more participatory and locally-owned policies like the PRSP and general budget support are one of the main reasons why they have lost favour in development agencies in recent years. One perverse consequence of the turn to results measurement is thus the tendency to replace more deliberative kinds of accountability with thinner more readily quantifiable forms.

## Conclusion

Faced with the growing complexity of the global environment and confronted with criticisms from borrowing governments and NGOs, development agencies have made important changes to how they design and deliver aid. In the process, they have introduced a number of policies that resemble the kinds of experimentalist initiatives that have become prevalent in the EU and in certain global policy areas: they have delegated more authority to debtor governments and civil society groups, developed participatory processes for identifying policy priorities, and introduced a range of new mechanisms to encourage horizontal accountability. Over time, however, the techniques of results-measurement that were supposed to guide policy efficiency and ensure accountability have come to dominate the process, with often-perverse consequences.

Although the power imbalances that define the relationship between donors and borrowers is far more pronounced in international development than in other areas of global or regional governance, the drive towards more measurement-driven forms of accountability and control is common to all of them. Many governments and international organizations have adopted results and performance-based approaches to their tasks as a way of trying exercise control over an increasingly uncertain policy terrain. As we seek to understand the shifting patterns of global governance more broadly, we therefore need to pay attention to the role of measurement.

As the experimentalist governance literature points out, many recent innovations in policy development and implementation have the potential to democratize governance practices by delegating authority to local actors and replacing vertical principal-agent dynamics with more horizontal forms of accountability. Yet, this study suggests that there is also a darker side to these new governance practices: the increasingly reliance on certain technologies of measurement—particularly those that require the demonstration of results—ultimately undermines the quality of the forms of deliberation involved and distorts many of the policies implemented. In too many instances, the measurement techniques that were supposed to be handmaidens to the policy process have instead become its masters.

---

1 On the increasing use of quantitative targets in international development, see: S. Fukuda-Parr and A. Yamin, "The Power of Numbers: A Critical Review of MDG Targets for Human Development and Human Rights," *Development*, 56 (2013): 58-65. For broader discussions of the politics of measurement and peer review in global governance, see: Davis, K. et al., eds. *Governance by Indicators: Global Power through Classification and Rankings*. Oxford: Oxford University Press, 2012; A. Cooley and J. Snyder, eds., *Ranking the World: Grading States as a Tool of Global Governance* (Cambridge: Cambridge University Press, 2015); T. Porter, "Making Serious Measures: Numerical Indices, Peer Review, and Transnational Actor-Networks," *Journal of International Relations and Development*, 15, no. 4 (2012): 532-57.

2 L. Hooghe and G. Marks, *Multi-level Governance and European Integration* (Lanham: Rowman & Littlefield, 2001); R. Keohane and D. Victor, "The Regime Complex for Climate Change," *Perspectives on Politics*, 9, no. 1 (2011): 7-23; G. Marks, L. Hooghe and K. Blank, "European Integration from the 1980s: State-centric v Multi-level Governance," *Journal of Common Market Studies*, 34, no. 3 (1996): 341-378; A. L. Newman, "Building Transnational Civil Liberties: Transgovernmental Entrepreneurs and the European Data Privacy Directive," *International Organization*, 62 no. 1 (2008): 103-130; A.-M. Slaughter, *A New World Order* (Princeton: Princeton University Press, 2004).

3 An earlier version of this article was initially prepared for a workshop on experimentalist governance organized by Robert Keohane, Charles Sabel, Jonathan Zeitlin and Gráinne de Búrca, November 7, 2014 at Brown University. Although I received very helpful feedback from them and the other participants at the workshop,

all views expressed and any errors or omissions are my own. For some of the major writings on experimentalist governance, see: G. De Búrca, R. Keohane and C. Sabel, "Global Experimentalist Governance," *British Journal of Political Science*, 44, no. 3 (2014): 477-86; S. Eckert and T. Borzel, "Experimentalist Governance: An Introduction', *Regulation & Governance*, 6, no. 3 (2012): 371-377; C. Sabel and J. Zeitlin, "Learning from Difference: The New Architecture of Experimentalist Governance in the EU," *European Law Journal*, 14, no. 3 (2008): 271-327; C. Sabel and J. Zeitlin, *Experimentalist Governance in the European Union: Towards a New Architecture* (Oxford: Oxford University Press, 2010).

4 On multi-level governance see: Hooghe and Marks, *Multi-level Governance;* Marks, Hooghe and Blank, "European Integration"; on networked governance, see: Newman, "Building Transnational Civil Liberties"; Slaughter, *New World Order*; on regime complexes, see: Keohane and Victor, "Regime Complex"; and on experimentalist governance, see: De Búrca, Keohane and Sabel, "Global"; Sabel and Zeitlin, "Learning from Difference"; Sabel and Zeitlin, *Experimentalist Governance*.

5 As noted above, there is however a growing body of research into the role of numbers in global governance, to which this article contributes, as well as a long tradition of scholarship on bureaucratic expertise and "audit culture," including: M. Barnett and M. Finnemore, *Rules for the World: International Organizations in Global Politics* (Ithaca: Cornell University Press, 2004); M. Power, "Making Things Auditable," *Accounting, Organizations and Society*, 21, no. 2/3 (1996): 289-315.

6 De Búrca, Keohane and Sabel, "Global"; Sabel and Zeitlin, "Learning from Difference"; Sabel and Zeitlin, *Experimentalist Governance*.

7 Sabel and Zeitlin, 'Learning from Difference"; Sabel and Zeitlin, *Experimentalist Governance.*

8 De Búrca, Keohane and Sabel, "Global," pp. 5-6; Sabel and Zeitlin, "Learning from Difference," pp. 276-8.

9 On the delegation of authority to lower-levels of units in EU governance, see: Sabel and Zeitlin, 'Learning from Difference."

10 Ibid.

11 Ibid, pp. 272-6.

12 De Búrca, Keohane and Sabel, "Global," p. 479; Sabel and Zeitlin, "Learning from Difference," pp. 278, 303-305.

13 On this dynamic in the experimentalist literature, see: Sabel and Zeitlin, "Learning from Difference," pp. 303-5.

14 J. Best, 'Bureaucratic Ambiguity', *Economy and Society*, 41, no. 1 (2012): 84-106.

15 Sabel and Zeitlin, "Learning from Difference," pp. 274, 324-5.

16 De Búrca, Keohane and Sabel, "Global," pp. 278-9; Sabel and Zeitlin, "Learning from Difference," p. 306.

[17] As Barnett and Duvall point out, just because compulsory forms of power are not obvious does not mean that subtler forms of power are not in play. M. Barnett and R. Duvall, "Power in International Politics," *International Organization* 59, no. 1 (2005): 39-75.

[18] I provide a much fuller account of these policies' failure (and the dynamics around policy failure more generally) in: J. Best, *Governing Failure: Provisional Expertise and the Transformation of Global Development Finance* (Cambridge: Cambridge University Press, 2014).

[19] UNDP, "Human Development Report 1999," New York: United Nations Development Programme, 1999.

[20] Success rates declined from 80 to 65 percent by the early 1990s. OED, "Annual Review of Evaluation Results 1993," Washington: Operation Evaluation Department, World Bank, 1994.

[21] Best, *Governing Failure,* Ch. 4.

[22] These external pressures for change were also combined with internal pressures for reform, particularly at the World Bank where a marginalized cadre of social science-trained staff saw their chance to finally bring the social and political dimensions of development back into the policy framework. Since they had to do so in terms that were recognizable to the still-dominant economists at the Bank, institutionalist economics ultimately became a crucial means for bringing (a particular version of) the social back in. Interviews with former and current senior World Bank staff members, May 10 and 11, 2007 and October 7, 2008, Washington D.C. See also: A. Vetterlein, "Economic Growth, Poverty Reduction, and the Role of Social Policies: The Evolution of the World Bank's Social Development Approach," *Global Governance*, 13, no. 4 (2007): 513-533; J. Best, "Redefining Poverty as Risk and Vulnerability: Shifting Strategies of Liberal Economic Governance," *Third World Quarterly*, 34, no. 1: 109-29

[23] OECD, "Paris Declaration on Aid Effectiveness," Paris: Organisation for Cooperation and Development, 2005.

[24] The literature on political ownership is vast, including: IMF, "Strengthening Country Ownership of Fund-Supported Programs," Washington, DC: International Monetary Fund, 2001; OECD, "Final Report of the Ad Hoc Working Group on Participatory Development and Good Governance," Paris: Organisation for Economic Cooperation and Development, 1997; World Bank, *Assessing Aid: What Works, What Doesn't, and Why* (New York: Oxford University Press, 1998).

[25] D. Dollar and J. Svensson, "What Explains the Success or Failure of Structural Adjustment Programs?," *World Bank Development Research Group.* Washington, DC: World Bank, 1998. This was also the conclusion reached by: P. Collier, "The Failure of Conditionality," in C. Gwyn and J. Nelson. (eds.) *Perspectives on Aid and Development* (Washington, DC.: ODC, 1997): 51-77; T. Killick, "Principals, Agents and the Failings of Conditionality," *Journal of International Development*, 9, no. 4 (1997): 483-494.

[26] Best, *Governing Failure*, Ch. 5.

[27] D. Warrener, "Current Thinking in the UK on General Budget Support," London: Overseas Development Institute, 2004; OECD-DAC, "DAC Guidelines and Reference Series: Harmonising Donor Practices for Effective Aid Delivery," Paris: Organization for Economic Cooperation and Development - Development Assistance Committee, 2006.

[28] OECD, "Evaluation of General Budget Support: Synthesis Report," Paris: Organisation for Economic Cooperation and Development, 2006, p. 1.

[29] Sabel and Zeitlin, "Learning from Difference," p. 280.

[30] World Bank, "What are PRSPs?" Washington, DC: World Bank Group, 2014. http://go.worldbank.org/CSTQBOF730. Retrieved August 27, 2014.

[31] G. De Búrca, 'Developing Democracy beyond the State', *Columbia Journal of Transnational Law*, 46, no. 2 (2008): 221-278; G. De Búrca, R. Keohane and C. Sabel, "New Modes of Pluralist Global Governance," *New York University Journal of International Law and Politics*, 45, no. 3 (2013): 723-86.

[32] IEO, "Evaluation of the IMF's Role in Poverty Reduction Strategy Papers and the Poverty Reduction and Growth Facility," Washington, DC: Independent Evaluation Office of the International Monetary Fund, 2004; N. Molenaers and R. Renard, "The World Bank, Participation and PRSP: The Bolivian Case Revisited," *The European Journal of Development Research*, 15, no. 2 (2003):133-161; L.-H. Piron and A. Evans, "Politics and the PRSP Approach: Synthesis Paper," London: Overseas Development Institute, 2004.

[33] De Búrca, "Developing Democracy," p. 276.

[34] One of the things that struck me in my interviews with aid agency staff between 2005 and 2012 was how quickly the emphasis shifted from ownership to results as the dominant framework for defining aid success.

[35] Practical Concepts, Inc. "The Logical Framework," Washington, DC: USAID, 1971; USAID, "The Logical Framework," in *Project Planning and Management Series* Washington, DC: United States Agency for International Development, 1980.

[36] D. Osborne and T. Gaebler, *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector* (Reading, MA: Addison-Wesley, 1992).

[37] R. Chambers, "The Origins and Practice of Participatory Rural Appraisal," *World Development*, 22, no. 7 (1994): 953-69. Interview with senior World Bank staff member, May 10, 2007.

[38] CIDA, "A Hitch-Hikers Guide to Managing for Results," Ottawa: Canadian International Development Agency, 1994; CIDA, "Draft CIDA Policy for Results-Based Management (RBM)," Ottawa: Canadian International Development Agency, 1995.

[39] DFID, "Results Action Plan," London: Department for International Development, 2007.

[40] Ibid.; CIDA, "2008 Results-Based Management Policy Statement," Gatineau: Canadian International Development Agency, 2008. Interview with UK Government official, February 15, 2010, London, and interview with David Booth, Senior Research Fellow, Overseas Development Institute, February 9, 2010.

[41] World Bank. "The World Bank and Aid Effectiveness: Performance to Date and Agenda Ahead." Washington, DC, 2011.

[42] World Bank. "World Bank for Results," Washington, DC, 2011.

[43] World Bank. "Program-for-Results Two-Year Review: Concept Note," Washington, DC, 2013.

[44] For my own analysis, see: Best, *Governing Failure*, Ch. 8.

[45] Interview with Ruth Driscoll, Research Fellow, Overseas Development Institute, September 11, 2006, London.

[46] MCC, "Guide to the MCC Indicators and the Selection Process for Fiscal Year 2014," Washington, DC: Millennium Challenge Corporation, 2013.

[47] However, since the MCC's projects have to comply with US procurement guidelines and management standards, it is difficult for the agency to rely much on country systems S. Lucas, "Millennium Challenge Corporation Principles into Practice: Country Ownership," Washington, DC: Millennium Challenge Corporation, 2011, pp. 3, 25.

[48] S. Lucas, "Millennium Challenge Corporation Principles into Practice: Focus on Results," Washington, DC: Millennium Challenge Corporation, 2011, pp. 14-15.

[49] Ibid. pp. 7-8.

[50] EC, "The Future Approach to EU Budget Support to Third Countries," Brussels: European Commission, 2011.

[51] O. Barder and N. Birdsall, "Payments for Progress: A Hands-Off Approach to Foreign Aid," *Center for Global Development Working Paper*, no. 102, 2006.

[52] DFID, "DFID's Approach to Payment by Results," London: UK Department for International Development, 2013.

[53] Barder and Birdsall, "Payments," p. 3.

[54] Ibid, p. 15.

[55] Interviews with senior World Bank staff members June 15 and 17, 2010, with IMF staff members June 14, 2010, and with OECD staff members May 2 and 4, 2011.

[56] Lucas, "Results," pp. 15, 17.

[57] Interview with OECD-DAC staff member, May 4, 2011, Paris.

[58] There are numerous cases recorded, particularly in the healthcare sector, in which targets have had perverse effects. A *Sunday Telegraph* report in the UK, for example, noted that hospitals were making patients wait outside or in corridors for hours at a time before admitting them in order to ensure that they met the government's target of treating them within four hours of admission. L. Donnelly, "Patients forced to wait hours in ambulances parked outside A&E departments," *The Telegraph*, May 30 2009.

[59] The "MCC Effect" is a term that the Millennium Challenge Corporation itself has begun to use on its website.

[60] Lucas, "Principles," p.15.

[61] EC, "Future," p. 3.

[62] Interview with Former World Bank Senior Staff Member, May 9, 2007, Washington, DC. These comments echo those I heard from several other World Bank and IMF staff members.