



USAID
FROM THE AMERICAN PEOPLE

EVIDENCE REVIEW PAPER

Responsiveness and Accountability in Local Governance and Service Delivery

An Agenda for USAID Program Design and Evaluation

May 2013

This publication was produced as part of the United States Agency for International Development's Center of Excellence on Democracy, Human Rights, and Governance Impact Evaluation Initiative. It was prepared by Jonathan Rodden, PhD (Stanford University) and Erik Wibbels, PhD (Duke University).

United States Agency for International Development
Center of Excellence on Democracy, Human Rights, and Governance

EVIDENCE REVIEW DRAFT PAPER

Responsiveness and Accountability in Local Governance and Service Delivery

An Agenda for USAID Program Design and Evaluation

May 2013

This draft Evidence Review Paper is one of several in a series commissioned by the USAID's Center of Excellence on Democracy, Human Rights and Governance. It was prepared by Jonathan Rodden, PhD (Stanford University) and Erik Wibbels, PhD (Duke University) through Social Impact. All rights reserved. The findings, interpretations and conclusions expressed in this paper are entirely those of the author(s) and should not be attributed in any manner whatsoever to USAID.

Table of Contents

Introduction	1
Section 1. Program Theory: An Analytical Framework for Thinking about Decentralization, Responsiveness and Accountability.....	3
1.1 Intellectual Background	3
1.2 Three Approaches to Improving Incentives.....	4
Section 2. Responsiveness	5
2.1 Existing Research.....	5
2.2 Potential Interventions and Evaluations	9
Section 3. Retrospective Accountability to Local Citizens	11
3.1 Existing Research.....	11
3.2 Potential Interventions and Evaluations	13
Section 4. Hierarchical Accountability.....	17
4.1 Existing Research.....	17
Section 5. Review of USAID Programming.....	19
5.1 Program Theory	20
5.2 Program Design	20
5.3 Program Evaluation	21
Section 6. Program and Evaluation Design.....	22
6.1 The Advantages of Randomization	22
6.2. Quasi-Experimental Alternatives to Randomization	23
6.3 Deciding Geographic Level of Interventions	24
6.4 What Outputs Should Be Measured?.....	25
Conclusion.....	26
References	28

Introduction

Global trends toward democratization and decentralization since the 1980s have not resolved the basic problems of governance in developing countries: corruption, low effort, inefficiency, and inequity in service provision. Moreover, ensuring that the most vulnerable populations within these countries have basic access to services has proven to be a monumental challenge. As highlighted by the World Bank Development Report (2004), the challenge is driven in large part by conflicts of interest that arise from the multiple principal-agent relationships inherent in public service delivery. These problems lead to glaring mismatches between what citizens want from government and what they get.

As new layers of government are created and lower-level politicians and interest groups are empowered, the attribution of responsibility for specific outcomes becomes less clear to citizens and donors. While responsiveness and accountability in public policy have traditionally centered on a relationship between the electorate and politicians, the agenda has also shifted increasingly to the behavior of bureaucrats and service providers. Frontline misconduct of public servants—termed “quiet corruption”—encompasses a wide set of transactions (i.e. doctor-patient, teacher-pupil interactions) that directly affect the intended beneficiaries of public services, often with detrimental long-term consequences (World Bank 2010). Efforts to improve responsiveness, accountability, and service delivery have thus emerged at the forefront of the policy agenda for USAID and many other development agencies.

While responsiveness and accountability are crucial concepts, they are often poorly defined and only loosely connected to specific programmatic interventions. In particular, there is little evidence to indicate what programmatic interventions are effective in promoting a better match between citizen needs and government policies. This paper disentangles three distinct approaches to intervention and explores the ways in which they can serve as a basis for policy design and evaluation

- 1) First, we identify a cluster of interventions aimed at enhancing the *responsiveness* of government to local citizens’ policy preferences.
- 2) Second, we examine interventions that seek to enhance *accountability* by empowering voters to punish and reward local officials for good or bad past performance.
- 3) Third, we examine interventions that rely on higher-level officials rather than local voters to fulfill this function.

There is still much to learn about governance and accountability, and scholars are increasingly turning to more rigorous evaluations to better understand the dynamics of decentralized decision-making and local governmental accountability (Alatas et al. 2010; Banerjee et al. 2010; Ferraz and Finan 2008, 2011; Olken 2007, 2010). Because of the scope and breadth of its programs to support democratic development, USAID will benefit greatly from its recent efforts to support impact evaluations and amass evidence-based knowledge on best practices and methods for program implementation. In doing so, it will contribute

to: 1) ongoing attempts to understand the fundamental ingredients of decentralized decision-making and government responsiveness to citizen needs and 2) improving the methods used to assess and effectively evaluate democracy and governance (DG) improvement programs.

This document is organized into five subsequent sections and sets out some basic analytical concepts around which programs related to responsiveness and accountability might be designed in the future as well as providing a roadmap for a research agenda of innovative program evaluations.

Section 1 introduces our program theory by describing in greater detail three distinct causal mechanisms through which decentralized governance may be brought into better alignment with citizen needs: 1) improved responsiveness to local citizens; 2) improved accountability to local citizens; and 3) and improved accountability to higher-level officials.

Sections 2 through 4 explore each of these mechanisms in turn, reviewing the existing theoretical and empirical literatures and then exploring avenues for a USAID evaluation agenda. Section 5 discusses existing USAID programming and evaluation efforts and recommends some areas for improvement. Finally, section 6 turns attention to some of the most crucial program design challenges facing USAID as it moves forward with an ambitious evaluation agenda.

After reviewing the existing theoretical and empirical literatures and then exploring avenues for a USAID evaluation agenda, we reach the following broad conclusions:

- An empirical literature based on rigorous impact evaluations of interventions related to responsiveness and accountability is still early in its development.
- There are no “silver bullets” that yield improved outcomes in every setting. It is not possible to make simple statements about “what works and what does not.”
- Rather, research is slowly identifying a set of conditions under which specific interventions yield improvements on rather specific outputs. The same intervention can yield rather different results under slightly different conditions.
- In the presence of mixed results thus far, the goal of academic research and the USAID learning agenda is to establish the precise conditions under which programs are most likely to succeed.
- It is not necessarily the case that “all good things go together.”
- It is possible to achieve “intermediate” results, like increased participation of civil society groups or individuals, without improving government policy or performance.
- Thus when designing programs, it is necessary to work hard to carefully articulate a program theory linking specific interventions to specific measurable outcomes.
- The evaluation agenda must be considered as early as possible in the process of program design in order to facilitate the clearest possible link between specific interventions and specific outcomes.

- This type of evaluation agenda will sometimes require simpler, less multifaceted program design.
- As research continues to sort out which interventions work under what conditions, USAID is in an excellent position to lead the way forward in a learning agenda that will generate a global knowledge pool about responsiveness, accountability, and good governance.

Section 1. Program Theory: An Analytical Framework for Thinking about Decentralization, Responsiveness and Accountability

1.1 Intellectual Background

Development donors and lenders in the 1980s and 1990s were optimistic about the trend toward decentralization that often accompanied democratization. Initially, support for decentralization was viewed as an end in and of itself. Based on treasured notions from welfare economics and public choice theory, it was taken as a given that decentralization would lead to better governance and service delivery.

By the end of the 20th century, however, this optimism was gone. Time and again, donors and lenders discovered that newly empowered local officials and bureaucrats did not face sufficient incentives to understand and accommodate the desires of broad groups of local citizens. Far from being a panacea, decentralization often created new opportunities for corruption and rent seeking that facilitated the capture of decentralized politics by local elites and interest groups. While a limited body of evidence suggests that the introduction of decentralized elections can promote public goods expenditures (Martinez-Bravo et al. 2012) and that citizens can potentially hold local public officials accountable informally through social groups (Tsai 2007), evidence overwhelmingly suggests the local elite capture can obstruct accountability (Olken 2007; Veron et al. 2006). It is now clear that embedding new strata of rent-seekers in a complex system of multilevel governance can actually attenuate the ability of citizens to identify and punish poorly performing governments.

As a result, the discussion in the development community has shifted from fostering decentralization to minimizing its dangers. This occurred in conjunction with a move toward the belief that “good governance” is the key to improving development outcomes. Indeed, “governance”—broadly understood as the relations among citizens, politicians, public bureaucracies, and the private sector—has become a fashionable term for academics and the international development community alike. Good governance is associated with lower levels of corruption, higher levels of bureaucratic professionalism, higher life expectancy, better water quality, lower poverty, and stronger economic growth (Holmberg et al. 2009).

Responsiveness and accountability have come to be seen as crucial ingredients of good governance. Indeed, the failures of decentralizing reforms are often now attributed to

weaknesses in the incentive structures facing local politicians and bureaucrats rather than simply a lack of capacity. Existing research provides very limited evidence that capacity building unto itself is the key to good governance, and it is for that reason that we (and most researchers) focus on incentives in the pages that follow. The key challenges are incentives that promote laziness, corruption, and other rent-seeking behavior that serve to distort democratic processes and obfuscate lines of accountability. In addition to outright theft and misuse of funds, academics and donors are also beginning to focus on other forms of poor performance, such as absenteeism of health workers and teachers, and underutilization of existing resources. Donors and lenders now believe that a crucial component of success for many of their efforts in decentralized settings is the construction of mechanisms of responsiveness and accountability.

1.2 Three Approaches to Improving Incentives

The initial enthusiasm among academics and practitioners about decentralization was based on two distinct notions: that it would enhance the match between preferences and policies, and that it would reduce agency costs. It is useful to develop this distinction further, for it delineates two basic approaches to understanding democratic representation, and hence two basic approaches to program design and evaluation.

In the first view, democracy is a tool for discovering the “will of the people” through election campaigns in which candidates offer, and then implement, distinctive platforms that reflect constituent preferences or demands. Democracy is valuable because it forces competitors to gather accurate information about citizen preferences and articulate these preferences through public policies in order to win. In this view, democracy performs well when there is a strong correspondence between measures of citizen preferences and government outputs. We might conclude that a decentralization program is working well if we can determine that a large majority of citizens in a community would rather invest in water pumps than new roads, and we observe that an elected government provides them with water pumps rather than roads.

This view of representation is sometimes referred to as “prospective” since it requires politicians to offer competing platforms about policies that they promise to enact in the future. In their influential textbook, Persson and Tabellini (2002) refer to this as “pre-election politics,” since campaigns and policy promises play a central role. We refer to this accountability mechanism as *responsiveness*.

In a classic argument, William Riker (1982) casts doubt on this understanding of democracy. Drawing on the work of Condorcet, Arrow, and other social choice theorists, Riker suggests that it rarely makes sense to think of the public as having a “collective will.” He offers a more minimal, retrospective view of democracy. When voting in elections, citizens are not expressing preferences over policies, but rather, making retrospective evaluations about whether or not politicians have performed well.

These retrospective evaluations are difficult because voters have very little reliable information about the choices facing politicians, and politicians have incentives to hide information in order to protect their own interests. Since information is scarce and people have limited time, voters often use information shortcuts based upon everyday experiences to make evaluations of economic performance, service provision, or property values to judge how their government is performing (Fiorina 1981, Popkin 1995). When these indicators fall below some threshold, voters throw elected officials out and give someone else the opportunity to do better. As long as politicians desire to retain power, this retrospective judgment can be an effective way of keeping politicians from stealing too many of the fruits of office for themselves or narrow interest groups (Besley 2006). We refer to this accountability mechanism as *retrospective accountability*.

These two clusters rely primarily on a direct link between voters and local politicians. However, local politicians must be able to induce good behavior on the part of unelected bureaucrats and service providers as well. Moreover, in many settings, local bureaucrats are agents of the central government rather than of the local voters, and in these cases, the best way to combat corruption and rent seeking might involve oversight and incentives from the central government.

In most of the developing world, local public services are funded through some form of intergovernmental transfers. This raises the possibility that without a strong local link between taxes and benefits, local citizens will have weak incentives to monitor local officials, since they are viewed as agents of a distant central government. Indeed, this perception is often accurate, since a part of the profile of local government responsibilities often involves administering policies that are “owned” by the central government. In these settings, the best hope for improving incentives, monitoring, and oversight might come from the center rather than local citizens. Thus, we refer to this accountability mechanism as *hierarchical, or top-down, accountability*.

To summarize, improvements in matching citizen preferences with government policies can be achieved through three mechanisms:

- *Responsiveness*. Improving the incentives for lower-level politicians and bureaucrats to learn about and respond to the needs of citizens,
- *Retrospective Accountability*. Improving the capacity of citizens to evaluate the past performance of elected and appointed officials, and
- *Hierarchical Accountability*. Improving the capacity of higher-level governments to monitor and sanction lower-level officials.

Section 2. Responsiveness

2.1 Existing Research

Part of the appeal of decentralized service delivery relies on the belief that citizens have coherent and identifiable preferences about services, and these vary from one locality to

another. Thus local policies can be tailored to the needs and desires of local populations more effectively than when decision-making is centralized. Schools can carry out instruction according to local language and tradition. Resources for public works can be funneled into the projects with the greatest local payoff.

It is in the spirit of enhancing responsiveness that USAID and other donors have set out to “enhance” or “broaden” local democracy. The central notion is that elite capture thrives when the electorate is disengaged and inactive, and policies can better approximate the “collective will” if interventions encourage electoral turnout, lobbying, attendance at meetings, participation in the budgeting process, and other forms of engagement with local government. Such initiatives range from increasing electoral participation to enhancing the capacity of NGOs to engage in participatory budgeting.

While greater participation can indeed be a good thing and can perhaps be achieved at relatively low cost (Gerber and Green 2008), there are many unknowns and some potential dangers. Indeed, some efforts to enhance the role of citizen groups in the policy process might invite greater involvement of interest groups, favoring those groups that can overcome the costs of collective action. Not only might this favor groups with more resources, but also small groups with high stakes (like public employees or firms who contract with local governments), who might find it easier to organize and pressure governments than larger groups (e.g. service users) with more diffuse interests. The interests of some community members in long-lasting and lucrative public employment opportunities might conflict with broader goals of efficient service provision. Generally speaking, there are many unanswered questions about the impact of participation on local service provision.

Nevertheless, existing research makes clear that the conditions under which participation takes place varies considerably across localities, and the nature of participation matters a great deal for the responsiveness of government. Below we discuss four key issues bearing on government responsiveness:

1. The prevalence of local clientelism
2. The heterogeneity of local communities and the level of participation across groups
3. The complexity of the issues that government deals with
4. The proliferation of sector-specific governing bodies

First, the political and economic organization of many developing countries is characterized by patron-client transactions (Stokes 2005; North et al. 2009). The clientelistic exchanges typical in many developing countries often involve citizens voting or otherwise participating in democratic procedures in a manner consistent with their patron’s wish in return for some private transfer. Vote buying, targeted transfers, and nepotism are characteristic examples (Finan and Schechter 2012; Wantchekon 2003, 2009; Olken 2007).

Where clientelism is pervasive, local elections and politics often have little to do with service provision, and elites are able to tailor budgeting and service provision to suit their

own needs. Instead of competing on the quality of public services, politics can focus on ethnic, linguistic, and regional rivalries, or the exchange of cash and small favors for votes. In response to this class of issues, Wantchekon (2003) convinced party leaders to eschew traditional clientelistic practices and focus on appeals related to service provision in constituency strongholds. He shows that clientelism is often effective *and* produces widespread electoral support. In subsequent work (2009), he attempts to bolster the credibility of programmatic platforms, with policies recommended by technical expertise, and finds that programmatic policy platforms can increase electoral participation and citizen knowledge. However, it is unclear if the treatment impacted the motivation or information (or both) of voters, and his interventions would be very difficult to replicate elsewhere. Nevertheless, the underlying implication of research on clientelism is clear: Increasing participation when clientelistic exchange is the norm is unlikely to improve the responsiveness of local government.

Second, as demonstrated by a generation of social choice theory from Condorcet to Arrow, it is not straightforward to characterize any group of individuals as having a meaningful “collective will.” It is possible that instead of reducing elite power and revealing the public’s true desires, greater participation merely introduces a more heterogeneous cacophony of voices and social conflict. Thus, one should not presume that widespread participation will lead to collectively optimal decisions.

A recent set of interventions explicitly recognizes the difficulty of characterizing a community’s “collective will,” and focuses instead on the fact that different groups have different preferences over service provision: women and men, old and young, rich and poor, high and low caste, etc. Because aggregating citizen preferences is difficult and methods of retrospective accountability are limited, principal-agent problems inevitably arise. That can have a substantial impact on the design of policies, and thus far, research has focused on the impact of rather blunt ways of favoring specific groups: initiatives to increase participation among particular groups (Chattopadhyay and Duflo 2004), and reserving seats or other forms of affirmative action for traditionally marginalized groups (Pande 2003; Beaman et al. 2010). These interventions often, but not always (i.e. Bardhan et al. 2010), find significant effects of representation on policies and citizen attitudes. In a related paper, Banerjee et al. (2010) report that a treatment aimed at decreasing voting along caste lines increased participation, although a related one aimed at corruption did not. The overarching message of this area of research is simply that *who* participates matters a lot, and increasing the number of participants does not necessarily make it easier to identify the public’s collective will.

Third, it is increasingly clear that the capacity of citizens to promote responsive government is constrained by their knowledge of what government does and the method through which they participate. A substantial body of work shows that local citizens are often confused about which level of government provides which services, and this can produce sharp breaks in accountability (Lyons et al 1992; Gelineau and Remmer 2006). A related problem, addressed in recent research by Khwaja (2004) is that many important local decisions about service provision are technical in nature, and substantial participation by non-experts can introduce inefficiencies.

As the complexity of issues grows, the precise nature of citizen input is likely to matter a great deal. For instance, Olken (2010) finds substantial differences between plebiscites and representation-based meetings, and a large body of observational research suggests that specific rules are extremely important in democracies: mechanisms for determining agenda control, open versus closed rules in legislative bodies, secret versus public balloting, and at-large versus districted representation.

Fourth, a related body of work focuses on the introduction of sector-specific consultative bodies. International donors often find it convenient to circumvent the existing apparatus of local government and create their own quasi-governmental consultative bodies in order to allow for citizen input and feedback about specific programs. Indeed, there is a compelling logic to encourage things like parental partnership with teachers and health care providers in the joint production of education and child health.

Yet a number of hard questions must be addressed about these bodies. Observational studies (Berry 2009) suggest that the proliferation of special purpose districts (a particular variant of sector-specific bodies) is associated with decreased citizen participation, capture by special interest groups associated with specific sectors, and common pool problems as each special district over-fishes the common pool of local revenue. More micro-level studies emphasize the difficulty of mobilizing citizen interest in these bodies. Banerjee et al. (2010) in one of a small number of studies that address this issue, find that concerted campaigns aimed at informing citizens about their rights to participate in consultative school committees have no impact on community involvement, teacher effort, or learning outcomes in Uttar Pradesh, India. Khwaja (forthcoming) reports more encouraging findings on the effect of community participation on infrastructure projects in Pakistan, but even there, the benefits declined as projects became more complex. A large and growing number of impact evaluations are aimed at this kind of programming, but the results are varied—with positive effects in some places, no effects in others, and even negative effects in a few cases.

From all of this work on responsiveness, we conclude that:

- More remains to be learned about the conditions under which government is sufficiently responsive to citizen preferences.
- Increasing participation, in the context of clientelism, is unlikely to improve responsiveness.
- Increasing participation may not produce a clearer definition of the “public’s collective will”.
- However, increasing the participation of specific groups (caste, gender, etc.) can improve the responsiveness of government to those groups.
- Citizen participation may be more useful as the issues under consideration are less technically complex.
- Enthusiasm for sector-specific governance bodies has greatly exceeded evidence on their capacity to improve governance.

2.2 Potential Interventions and Evaluations

USAID is in an excellent position to build variation into DG programming that is designed to affect the mechanisms through which the preferences of voters and interest groups are filtered through the local policy process, such that its effects can be rigorously identified. Some of the areas of greatest opportunity are in broad decentralization reforms, where USAID might be able to work with governments to better understand the outcomes of different decision-making structures. Through a coherent research agenda, USAID also has the ability to test the external validity of the emerging findings by conducting similarly designed evaluations in different contexts.

Drawing on existing research and USAID programming, we discuss some specific examples of potential interventions and associated evaluation designs:

1) Encourage participation. A simple hypothesis is that governments only reflect the preferences of those who vote. Thus, interventions that broaden the group of voters might generate a mix of public services that is more responsive to the preferences of groups that previously did not vote. Such programming would be particularly promising in places where local citizens incorrectly perceive local elected officials to be mere pawns of the central government or where local officials have extensive formal, but as of yet underutilized, authority. A potential evaluation design could consist of first, administering a baseline survey to gather information about voting behavior and preferences over public goods in both treatment and control districts, disaggregating by social status, gender, etc. Next, implementing activities in treatment districts that would directly encourage voting (e.g. get-out-the-vote campaigns, information about elections, polling places, the authority of local officials, etc.). Evaluators could then examine whether, after some period of time, 1) turnout increased, and if so, among which social groups, 2) the mix of public goods changed to reflect the preferences of groups experiencing increased turnout, and 3) there is an impact on satisfaction with government.

2) Create new venues for direct citizen involvement. Another hypothesis is that in small communities, public services will better reflect the preferences of citizens if they are directly involved in decision-making through such procedures as participatory budgeting. Again, after a baseline survey is conducted as described above, introduce some new class of participation mechanism in treatment districts. After some period of time, evaluators can examine intermediate variables like participation of different groups, as well as service provision outputs and citizen satisfaction with participatory procedures.

2a) With sufficiently large budget and geographic scope, it may be possible to craft different participatory institutions in different districts. For instance, it would be useful to experiment with the selection mechanism for who is called to participate, as well as the institutional procedures through which decisions are made.

2b) Another variation has to do with sector-specific participation forums. For instance, participatory mechanisms might be built into local decision-making about infrastructure, health, or education policy. Following up on some of the insights from the research reviewed above, it may be possible to experiment with different ways of including experts and non-experts in such bodies and managing the complexity of the issues involved.

2c) A related question has to do with sector-specific versus catch-all participatory bodies. It would be useful to divide the treatment districts into two groups: one group is treated with a single catch-all participatory body aimed at something like overall public financial management, and the other group is treated with several sector-specific bodies. Then, the outcomes for both these treated sub-groups are compared to each other and to a control group.

3) Directly encourage representation of under-represented groups. In settings where women or certain social groups have traditionally been excluded from participation in civic affairs, it may be useful to build on the research reviewed above by experimenting with reserved seats or other mechanisms to directly boost the participation of these groups. After obtaining information about preferred public goods of different groups in the baseline survey, evaluators could then compare the mix of public services before and after the intervention in treatment and control municipalities.

4) Promote the secrecy of the ballot in decentralized elections. Research suggests that the secrecy of the ballot has important implications for the nature and level of citizen participation in politics (Gerber et al. 2008, 2012). In conjunction, recent work on clientelism emphasizes the creative ways in which local officials can violate ballot secrecy in order to monitor clientelistic electoral exchanges (Stokes 2005). If citizens are unable to freely express their views at the ballot box, it may deter participation and will certainly impact the message that local officials extract from electoral results. Interventions could aim at introducing electoral technology that discourages tampering or encourage tighter oversight of electoral procedures with the goal of deterring clientelism. Evaluators could then compare the mix of services and citizen satisfaction with democratic procedures in control and treatment localities.

Interventions and evaluations like these can help validate some of the existing findings in the literature, and continue with the gradual process of refining our knowledge about the specific conditions under which interventions are helpful. USAID has a distinct comparative advantage in that it has the geographic reach and requisite resources to experiment with different forms of participation across villages, districts and regions within countries and across countries.

Section 3. Retrospective Accountability to Local Citizens

3.1 Existing Research

Once government acts, citizens must have some means of evaluating and sanctioning it expost. This is the basic principle that motivates concern with retrospective accountability. Here, the democratic process is understood in terms of citizens making retrospective judgments about whether or not politicians have acted in congruence with the public interest. Citizens are viewed as the principals in a principal-agent relationship (Besley 2006).

The basic problem facing voters is that they often lack access to information. Furthermore, politicians often have incentives to hide information about their true intentions and level of effort. As Fiorina (1990) highlights, it is not clear whether it is rational for citizens to become informed at all, given that information acquisition is costly. Access to such information is especially difficult in developing countries because of inadequate infrastructure, government restrictions, low levels of literacy, and weak local presses. Indeed, it should be noted that a good deal of existing research raises serious questions about voters' ability to use information to assess politicians' performance. For instance, in the United States, voters apparently punish politicians for events beyond their control such as shark attacks, commodity prices, and the performance of hometown athletic teams (Bartels and Achen 2004a, 2004b; Healy et al. 2012). They are also sometimes deaf to information about past performance, placing undue weight on performance during a brief period right before elections (Huber et al. 2011). This "recency effect" can provide politicians with incentives to shirk during non-election periods and undertake a flurry of (potentially inefficient) activity in the period immediately before elections in political business cycles (Tufte 1978).

These generic problems of retrospective accountability can be exaggerated in contexts of decentralized governance. Observational research suggests that local voters often do not know which level of government provides them with which services. This problem is especially acute in countries where there are multiple layers of decentralized government, when responsibilities are being transitioned between levels of government (as in contemporary Ghana), or when new layers of government are being added whole cloth (as with the counties in Kenya or the regional tier now being considered in Uganda).

A related problem is with the politically motivated proliferation of governments within one tier, as with the explosion of the number of districts in Uganda, Sierra Leone, Ghana, and Indonesia. Finally, the nature of local public finance in many developing countries creates stark information problems for citizens. The vast majority of funds come from non-transparent intergovernmental grants, and surveys reveal that voters have very little understanding of where government revenue comes from, and do not see themselves as taxpayers with rights and responsibilities to monitor government activities. Local voters have no idea what resources are received by local officials, and for what purposes, making it difficult to hold them accountable for what they do with the money (see Paler 2012).

In light of these challenges, interventions in this area typically focus on various technologies for improving citizen information, including score cards, media campaigns, information sheets, comparative data on different officials or localities, and published audits. Banerjee et al. (2010), Humphreys and Weinstein (2010), and Grossman (in progress) for instance, report results of treatments in which citizens are provided report cards specifically on the behavior of legislators. Ferraz and Finan (2008) show that federal audits of randomly selected municipal governments facilitate electoral punishment as the number of reported violations increased. They also show that the presence of a local radio station to broadcast the results of audits sharply increased the electoral punishment of violating politicians. All told, the evidence is mixed: Citizens seem to hold politicians accountable for corruption that is revealed by audits, but they seem to care less about how their politicians spend their time.

Beyond efforts to impact the capacity of citizens to hold politicians accountable, other interventions aim to provide citizens with better information on the behavior of bureaucracies and service providers. Even the best-intentioned politicians are constrained by an unresponsive bureaucracy. In many settings, poor public services arise from worker absenteeism, lack of effort, and weak incentives rather than the behavior of politicians themselves. This has inspired some recent research that specifically targets local service providers. Andrabi et al. (2009), for instance, report on an intervention that provides details on school performance, which has measureable effects on both learning and school fees. Duflo et al. (2007) and Muralidharan and Sundaraaraman (2007) target service providers directly and show that even a moderate improvement in incentives can induce improvements in teacher performance, although there are a host of studies showing it can be incredibly difficult to motivate local bureaucracies to change. Bjorkman and Svensson (2009) examined the effects of meetings that introduced a community-based monitoring system involving local health clinics in Uganda, and report impressive effects on service provision, utilization, and health outcomes.

The results in this literature are not uniformly positive. Keefer and Khemani (2011) examine the effect of radio-based campaigns on the importance of education and health services for citizens. They find that such campaigns increase the salience of such services and may increase citizen attention to how they are delivered, but they do not increase collective action or impact the quality of service delivery.

From all of this work on retrospective accountability to citizens, we conclude that:

- Citizens often lack crucial information on the outputs of decentralized government, and this limits their capacity to hold governments accountable for their performance.
- Evidence on different approaches to resolving these information shortfalls are mixed, but some findings are emerging:
 - Report cards on the behavior of politicians seem to have little effect thus far.

- Audits that uncover corruption by local officials can lead to improved accountability to voters.
- Report cards on the quality of services are sometimes effective but not always. It is unclear which services truly motivate citizens to hold government accountable.
- There is evidence that community-based monitoring of service providers can yield impressive results, but there is much to learn about the conditions under which such monitoring works best.

3.2 Potential Interventions and Evaluations

Once again, USAID is in a very strong position to use its interventions and associated evaluations to validate existing findings and explore the precise conditions under which various strategies might improve outcomes.

It is important to be as clear as possible in program documents about how interventions bearing on retrospective accountability are expected to work. For instance, interventions can aim at providing information about the behavior of politicians, the behavior of bureaucrats, or the quality of one or several public services. Each of these interventions implies a distinct, substantively important mechanism for promoting accountability.

USAID programming can strengthen retrospective accountability by helping citizens collect the relevant information in a variety of ways. Again, this is not meant to be an exhaustive list of possible interventions, and it is important to ensure that interventions are not muddled together.

1) Provide information about specific elected politicians. After conducting a baseline survey to ascertain citizens' perceptions about the performance of officials, service provision, preferences over public goods, etc., institute an information campaign in treatment districts aimed at providing specific information about the activities of elected officials. Building on the "scorecard" literature, this can include measures of effort, information from audits, information about use of funds, etc. The theory is that politicians will work harder and engage in less rent-seeking behavior because they fear electoral repercussions of negative information revelation. After some period of time, evaluators can then examine whether efforts of politicians have improved in the treatment districts, whether the mix of public goods is more in line with preferences, whether service provision has improved, whether losses associated with corruption have been reduced, and whether citizen satisfaction has improved.

2) Provide information about the quality of local government. As discussed above, it is not clear whether it is better to focus on individual politicians or on the local government as a whole.

2a) It may be valuable to develop a nation-wide system of ratings for local governments, allowing voters to compare their government with others. Ratings

and supporting information would be released in treatment districts. Ideally, the information would allow for easy comparisons with other governments. The theory is that politicians in treatment districts will work harder because they fear the electoral repercussions of negative information.¹

2b) Rather than providing catch-all ratings, it might be useful to release some raw (but contextualized) data. For instance, some countries have placed considerable resources into audits of local governments, and the information from these reports might be released through radio campaigns, bulletin boards, or local meetings. Alternatively, in highly transfer-dependent settings, it may be useful to pursue information campaigns aimed at lifting the cloak of secrecy from intergovernmental transfers. When reliable data are available, provide local citizens with information on the amount and timing of transfers received by sector, provide some indication of the intended purpose of the allocated funds, and provide information about how the funds were actually spent. The program theory is that corruption and laziness thrive in settings where voters have no information about revenue. When voters can follow the funds by reading a report posted on the local bulletin boards, they are in a position to ask hard questions about discrepancies.

3) Provide information and/or monitoring technologies for local service providers. As discussed above, a separate class of interventions is aimed at local service providers rather than elected politicians. Within a particular sector like health or education, obtain detailed information about service provision at the lowest level possible, e.g. individual health posts or schools. Perhaps using geographic information systems to obtain a sample of users, institute a baseline survey to measure attitudes and experiences toward service provision. After some period of time, examine whether changes have taken place in the treatment localities in attitudes toward service provision, satisfaction with government, actual service provision, and actual aggregate indicators like student achievement, literacy, and health.

3a) In the treatment areas, provide detailed information about specific service outcomes like absentee rates or drug stock-outs. These data will be most useful if placed in comparison with other schools or health clinics, or compared with national standards. As in the discussion of intergovernmental grants above, it may be useful simply to inform citizens about the resources that have been received by the schools or clinics, and the formal expectations associated with those resource transfers regarding operating hours, provision of basic medications, and teacher effort. The theory is that once citizens understand what resources have been provided and what is expected, they have a benchmark against which to compare local service providers. They can then pressure service providers directly, or enlist

¹ A potential problem with this approach is that the good overall reputation of the local council is a kind of collective good for the councilors, and each councilor may have incentives to free ride. Thus an intervention like this is most likely to work in councils with a strong and perhaps even directly elected executive (e.g. mayor).

elected local officials, MPs, higher-level officials, or NGOs to put pressure on local service providers (more on this below).

3b) Combine information with monitoring technologies. This is an area with considerable room for innovation and learning. Banerjee, Deaton, and Duflo (2004) experimented with paid monitors, but had little success. Following Bjorkman and Svensson (2009), community meetings might be convened to facilitate sustainable community-driven monitoring programs. Another possibility involves harnessing SMS technology so that citizens can use cell phones to report absenteeism, wait times at clinics, and availability of medications.

4) Experiment with politicians versus bureaucrats as the locus of accountability. Some researchers have focused on elected officials as the locus of accountability, and others have focused on local service providers. It would be useful to design programs that help ascertain which approach is more fruitful. As mentioned above, the existing research that focuses on local service providers is not very clear about the exact mechanism through which enhanced community monitoring of providers is expected to yield better service outcomes. Are citizens expected to exert direct social pressure on service providers, ostracizing them in the community if they fail to live up to expectations? Or are citizens expected to pressure local elected officials or national MPs, who in turn are expected to pressure bureaucrats? Finally, are citizens expected to bring their complaints directly to the relevant ministries, who then are expected to pressure poorly performing local officials?

We know very little about which of these mechanisms might be most effective. USAID programs might incorporate subtly different treatments, such that some communities are prompted to send text messages to ministries, and are told in a community meeting that local service providers are agents of the central government. Another set of communities might have a different type of community meeting focusing on the introduction of elected officials as advocates, and a different SMS system would be set up to provide information directly to elected officials. A third group of communities might have a meeting that focuses on the local community as directly responsible for monitoring and pressuring local service providers, with no mention of central officials and no SMS system.

5) Experiment with taxation and the tax-benefit link. There is a relatively clear theory in the literature in which reliance on local taxation rather than transfers is expected to yield better information, oversight, and ultimately accountability. Indeed, a common claim is that accountability is a function of the nature of the link between taxes and benefits. The theory is that voters will pay greater attention to government's behavior when they are paying direct taxes. They will ask questions, gather information, and put pressure on officials to spend their taxes wisely. They will also have more information about government revenue than in the case of pure transfer-dependence. The argument is clearest for broad-based taxes like household taxes or property taxes, but even narrow taxes, e.g. motorcycle license fees, might generate a small but vocal group of engaged and informed taxpayers. USAID programs might help test the link between taxes and accountability by phasing in programs that enhance local revenue mobilization.

5a) Experiment with techniques for building the local tax base. A major problem with autonomous local taxation in poor countries is the fact that there is a very limited tax base in poor rural areas. However, it is usually possible to levy minimal property or agricultural taxes, market fees, or vehicle fees. The main problem is that local politicians are extremely resistant to levying such taxes and fees even when they have the constitutional and statutory authority to do so, simply because they are politically unpopular. In other cases, the failure to tax might simply reflect the lack of detailed local cadastral surveys.

In a subset of treatment localities, a USAID program might offer matching grants or other fiscal rewards for local governments that reach certain revenue targets, potentially in collaboration with a central government that is motivated to enhance local revenue mobilization.² It would be important to conduct an information campaign so that voters are aware of the potential payoff from enhanced local tax collection. When piloting a new program, it may also be possible to introduce local taxation or user fees in some communities while funding the project purely through intergovernmental grants in others.

5b) Another intervention would be to directly link the new taxes with specific public expenditures in some subset of localities. For example, as part of the introduction of the new or enhanced taxation, public meetings might be held in order to set priorities or vote on specific public goods that will be funded with the new resources. It would be useful to contrast actual revenue collection in the treatment group that received the explicit linking of taxes and benefits with another treatment group that received only the new tax efforts. And it would be useful to track changes in perceptions of government and service provision, along with actual service provision outcomes, in both treatment groups as well as the control group.

Of course the goal is not merely to raise revenue, but to enhance accountability by providing stronger incentives for citizens to monitor local governments. Before and after implementing an intervention aimed at enhancing local revenue mobilization and reducing reliance on intergovernmental grants, a comparison of baseline and endline surveys in treatment and control groups might look for changes in attitudes toward government, taxation, service provision, and monitoring activity. Researchers might also develop “harder” measures of monitoring and engagement with local government. Perhaps they might set up SMS or other reporting systems or track participation in community meetings or communications with elected officials.

By helping to design and implement the next generation of programming in this area, USAID is in a good position to help answer some very important questions: Which types of information most motivate citizens? Do they care about social aggregates, like graduation

² One can look to the AFD, CIDA, DANIDA, and KFW financed District Development Facility in Ghana as an example, though the link between grants and district performance could be better designed.

rates, or the specific outcomes in their local schools and health clinics? How important is it to provide sector-specific cues about things like funding sources, higher-level regulation, and the tools at the disposal of local elected officials? Is it better to provide very specific information about who is responsible for what, or is it better to simply provide aggregate data on service provision and let voters sort it out? Do politicians fail to tax local sources because they simply lack the necessary cadastral information or because it is politically unpopular? What are the conditions under which increased local revenues promote better services and greater accountability?

Section 4. Hierarchical Accountability

4.1 Existing Research

Local governments are often funded largely through intergovernmental grants or shared taxes rather than direct own-source taxation. Moreover, local governments are often merely conduits for implementing programs under the direction of a strong, central government. Local elections are often window-dressing, and in spite of what the laws or constitution might say, authority is in the hands of the executive or their local representatives. Sometimes local governments only collect information about eligibility for central programs or help register participants in such programs. Often the primary activity of local governments is to maximize the size of revenue flows from the central government, and the best qualification for leadership is a set of partisan, ethnic, or informal connections that will help facilitate this.

When local governments play such roles, local democratic accountability is not the most obvious way to induce good performance. Local officials face incentives established by their national sponsors, and voters may prefer politicians who seem to be successful in attracting funds, even if those funds are spent in inefficient ways. Worse, in such cases where there is competition with other localities or ethnic groups over scarce resources, ruthlessness and venality may be qualities valued by voters.

As initial enthusiasm for decentralization has waned, so has the insistence that accountability of local officials and service providers must flow exclusively from the bottom up. Recent years have seen considerable investment in alternative accountability mechanisms that involve higher-level governments and independent agencies. Although bottom-up accountability of local politicians to voters is desirable, this might be difficult to achieve in settings where local officials are not elected, or where local politicians or service providers face few career incentives to worry about the concerns of local voters. In these situations, the best way to improve performance in the short or medium term might be to strengthen mechanisms of hierarchical oversight.

Chief among these is the independent audit. Several recent papers suggest that audits can be extremely effective. In Olken's (2007) field experiment in Indonesia, audits were found to be far more effective in reducing corruption than increased grassroots monitoring. Ferraz and Finan's (2008) work suggests that audits can strengthen retrospective

accountability by opening the books of local governments and providing a context for the information that is revealed. Both papers provide evidence of strong effects, but they leave several unanswered questions, including: Is better behavior motivated by the ex ante threat of audit or the ex post threat of a public newly informed about malfeasance? What probability of audit is high enough to elicit better behavior?³ What level of auditing is sustainable by the national government?

A body of evidence suggests that when national parties or officials have control over local nominations and/or campaign finance, local officials become more responsive to the national government (Rodden and Wibbels 2002; Jones, Saiegh, and Spiller 2002). While these levers seem impervious to foreign aid programming, they do point to the importance of understanding the central government's incentives when considering mechanisms to promote hierarchical accountability. Hierarchical mechanisms are only likely to help enhance accountability and service provision in instances where the executive, or some important and powerful players at the center, are motivated to improve local outcomes. This is most likely to be the case when electoral competition at the center is vigorous, free, and fair. In this setting, incumbent legislators, presidents, or prime ministers may face incentives to promulgate reforms that bring about improvements in local service provision.

From this nascent work on hierarchical accountability, we conclude that:

- The empirical literature on this mechanism is less developed than the others.
- Audits by central authorities can improve accountability, although key questions remain.
- Any effort to augment hierarchical accountability should take into consideration the interests and incentives of central officials.

4.2 Potential Interventions and Evaluations

In the appropriate settings, USAID can partner with the center in efforts at improving hierarchical accountability mechanisms in a variety of ways.

1) Experiment with audits. The empirical estimation of the impact of audits can be tricky, since the theory is that the fear of audits will induce better performance among the vast majority of audit-eligible localities that are *not* audited. While working with a central government to build a stronger auditing agency, USAID and other donors can randomize the roll-out of treatment or conduct a lottery in selecting which localities to target for the possibility of audits. As in the design described above, survey data and service provision data can be collected in treatment and control communities before and after the rollout of the treatment. The hypothesis is that perceptions of service delivery and satisfaction will

³ In the Indonesia project, treatment villages had a 100 percent chance of receiving an audit and effects were entirely through the pre-audit incentives of local officials. In the Brazilian case, 1 out of 211 municipalities was audited, and the effects were entirely through the information campaign and not through the pre-audit incentives.

improve, along with the quality of actual service delivery, in the communities that were *eligible* for audits.

When the number of districts and the budget are large enough, such interventions can be paired in some set of local governments with ex post information campaigns aimed at publicizing audit outcomes to local citizens. If properly designed, this would allow program evaluators to examine the ex ante and ex post effect of audits on the behavior of local officials.

2) Improving data collection and top-down monitoring. In many settings, central or regional government officials would very much like to combat problems like poor performance and absenteeism among local health workers and teachers, but they do not have the information or tools to gather information or provide punishment and rewards. In such settings, USAID programs can help build capacity for central government ministries to gather their own data about local service provision rather than relying on self-reporting of local governments. These tasks might also be fruitfully outsourced to contractors, consultants, or engineers to provide independent estimates of project costs, needs, and other technical information. The development of performance auditing capacity of local governments seems particularly promising, since audits of financial accounts cannot uncover whether local procurement choices are efficient or in the public's interest.

Any such USAID effort at central government capacity building or data collection might be phased in over time, such that for an initial period of years, only certain localities are subjected to the new regime. As with the designs above, survey responses from the baseline and endline surveys as well as innovations in observed service delivery and even health and education outcomes could be contrasted in the treatment and control groups.

3) Introducing non-discretionary punishments and rewards. When USAID is able to work with a motivated central government agency or ministry and collect high-quality information about local performance, it may be able to help craft non-discretionary ways of making the pay and benefits of local administrators contingent on performance. Salary might be linked to performance, and performance bonuses might be distributed to high-performing schools or clinics. Again, such programs could be phased in over time to facilitate comparison of eligible and non-eligible communities.

Section 5. Review of USAID Programming

While preparing a new agenda for learning and evaluation related to decentralized governance and accountability, it is useful to learn from what has been done in the past. We reviewed USAID programming based on three sources: First, the NAS' *Improving Democracy Assistance*; second, our detailed reading of documents for current and forthcoming programming in Zimbabwe, Uganda, Russia, Kenya, and Ghana; and third, 25 randomly selected USAID programs under the heading of "Public Sector Executive Function" and "Local Government and Decentralization." Even with concerted help from USAID staff and a team of research assistants, we were unable to gather detailed program

documents (i.e. Scopes of Work, PADs, program evaluations, etc.) on these 25 programs. In lieu of actual program documents, we did a content analysis of publicly available material culled from country profiles on USAID's website and mission websites.

In general, our review identifies room for improvement in the areas of program theory, design, and evaluation.

5.1 Program Theory

We found that the majority of programs lacked clear program theory that would link interventions to specific program goals. Accountability is associated with a mix of transparent procedures, high quality service delivery, citizen satisfaction, and other desirables. Sometimes it is not clear whether the underlying motivation for an intervention is a mechanism related to prospective responsiveness or retrospective accountability. Sometimes the goal of the project seems to be the empowerment of civil society groups or the enhancement of participation rather than improved public policy or service provision.

That said, projects do generally follow a set of overarching themes. In many programs, accountability is either implicitly or explicitly understood as a principal-agent relationship running from voters to local elected officials to local service providers. An important program goal is to tighten up one or more of these relationships such that the agent has less scope for acting contrary to the interests of the principal. In the background, there is also a principal-agent relationship running from voters to central officials, and then downward from those officials to local governments, sometimes via one or more middle tiers of government. Some projects also attempt to strengthen this relationship.

The principal-agent approach is useful, but programs and interventions are typically unclear about: a) Which principal-agent relationship they aim to impact; and b) How exactly they aim to affect outcomes. For instance, if an intervention provides information to local citizens about local service provision relative to national standards, what are they expected to do with it? Are they expected to provide direct pressure to local bureaucrats, local elected officials, national politicians, or all of the above?

It was admittedly hard for us to gather information on the detailed program theories for each project, but it is difficult to escape the impression that "governance" and "accountability" have been viewed as rather loose, catch-all concepts. That USAID does not have a clearing house for gathering, processing and analyzing programs makes organizational learning, as a whole, extremely difficult and prevents the program theory for decentralization and governance from moving forward.

5.2 Program Design

In the absence of clear program theory, many programs are characterized by a relatively large number of interventions rolled out at more or less the same time across multiple

localities, districts, or regions. Most of the programs we reviewed were extremely broad and ambitious, but this often resulted in a proliferation of interventions that were weakly integrated and designed in ways that make them difficult to evaluate. They encouraged implementers to try a wide variety of activities, usually including some mixture of grants to civil society groups, training seminars, technical assistance, information campaigns, enhanced audits, and funding to promote local user groups. The lack of sharp focus in programming is perfectly understandable given that the basic problem is multifaceted, but it does not promote an evidence-based learning agenda.

Program documents often adopt the attitude that improvements in accountability require progress on a variety of fronts. A typical approach is to choose some small number of districts or counties and treat them with a mixture of information campaigns, meetings, grants, and technical assistance. Evaluation then consists of contrasting some aggregate data on service provision or citizen satisfaction in the treated districts with the rest of the country. Even if the treated districts appear to experience improved outcomes, however, it is difficult to know which of the many interventions may have been responsible or what synergies, if any, exist between the bundled interventions. Moreover, the small number of treated districts or counties leaves the evaluation team with no statistical power and a non-random selection of the treatment group.

5.3 Program Evaluation

It seems that many programs are not evaluated, and among those that are, many of the evaluations are carried out by the contractors who implemented the program. Absent specific and systematic program theory, program design and program evaluation, discretion in the field can result in ad hoc programs and little or no capacity to evaluate their success or failure. Indeed, the ambition and breadth of current programming often make rigorous evaluation all but impossible.

A common reply to the critique of multi-faceted programming is that modest, targeted interventions cannot work unless they are combined with a set of broader accompanying interventions that are rolled out simultaneously. Unfortunately the evidence for such assertions is lacking, in large part because appropriate impact evaluations have not yet been done. In fact such hypotheses present opportunities for interesting research. When the hypothesis involves interactions between different interventions, an evaluation team can help design an intervention with several branches, as described in further detail below.

In conclusion, our review of USAID programming suggests that:

- The definition of accountability is often vague, and it is often unclear who is expected to be held accountable to whom.
- It is sometimes not clear by what mechanism accountability is to be achieved—through an improvement in information, democratic procedures or institutions, the quality of local bureaucracies, or mechanisms of oversight.

- Programs are often designed to maximize the number of interventions, but this makes the efficacy of any particular intervention difficult to assess.

Section 6. Program and Evaluation Design

Following from these lessons, we now turn to some more specific program and evaluation design challenges. In order for USAID and the academic community to make progress, interventions need to be designed and implemented in a manner to maximize their evaluability. Rigorous program designs rooted in a clear theory of accountability are a precursor to learning what interventions are most effective in promoting responsive and accountable governance. As USAID moves forward with an evaluation agenda based on some of the causal logic and potential interventions laid out above, it will be necessary to address some difficult challenges. Rigorous impact evaluations that adhere to USAID's Evaluation Policy (2011), including experimental and quasi-experimental design, allow for exploration of causality and attribution of outcomes of interest to USAID programming. This section briefly overviews best practices bearing on program and evaluation design.

6.1 *The Advantages of Randomization*

The gold standard for program evaluation is randomized control trials. Causal inference can be undermined when communities are assigned to treatment and control conditions for political reasons, because they are in “high priority” regions, or because they are perceived to be either higher performing or in greater need. Non-compliance with treatment or spillover from treatment communities can also undermine causal inference.

There are several advantages to building randomized evaluation into RFPs and program designs. First and most importantly, an experimental design offers the potential for USAID, other practitioners, and the academic community to learn which interventions succeed or fail and to identify the mechanisms through which they do so. Randomized assignment of a single intervention to a treatment group as described above would preclude them from being mashed together in incoherent ways and would maximize the capacity of USAID to know which interventions work.

Second, in many cases randomization will allow USAID to not only assess *if* a program worked, but also *how much* it worked. Understanding whether outcomes improved substantially or marginally is crucial to assessing the cost-effectiveness of interventions.

Third, randomized evaluation can be quite cost-effective, adding little to program budgets even as it offers the potential for learning. Esther Duflo and colleagues (2006: 3) summarize the advantages in this way, “Unlike the early ‘social experiments’ conducted in the United States-- with their large budgets, large teams, and complex implementations-- many of the randomized evaluations that have been conducted in recent years in developing countries have had fairly small budgets, making them affordable for development economists. Working with local partners on a smaller scale has also given

more flexibility to researchers, who can often influence program design. As a result, randomized evaluation has become a powerful research tool.”

6.2. Quasi-Experimental Alternatives to Randomization

The advantages of randomization are clear and broadly accepted by academics and evaluation professionals, but practical and political hurdles to randomization are common. Politicians are loath to give up the ability to target programs to their favorite localities, and USAID often faces significant constraints on geographic targeting associated with past and current programs. Program designers might also worry about the ethics of randomly choosing which communities will receive a potentially valuable intervention.

Some of these concerns can be alleviated with a simple phased implementation strategy. It is often possible (and practical) to implement the program first in a randomly selected subset of localities, with the plan of expanding to the full set after sufficient time has passed. Phasing in different interventions over time in a locality offers some leverage in identifying program effects. If it is not possible to change the USAID proclivity for complex, multi-faceted programs, an alternative is to locate individual interventions that can be rolled out in a way that circumvents contamination by the other simultaneous treatments. For example, if one set of complex interlocking interventions will take place at the regional level in half of a country’s regions, it might be possible to implement some smaller, more targeted program in randomly selected localities within the focus regions so that the broader region-level treatment is held constant.

In settings where randomization is not feasible, a rigorous quasi-experimental evaluation may still be achievable. Possibilities include regression discontinuity designs and various efforts to match the localities that have received the program with those that have not.

First, a regression discontinuity design might be feasible if a program is allocated to localities according to some objective indicators. This often creates an objective threshold that determines whether a community qualifies for the program or not. For example, a program might only be allocated to a community above or below a certain population threshold. Researchers can then examine a set of communities that fall just above and just below the threshold, with the assumption that there are no systematic differences between the communities just above and just below the threshold.

However, program designers should be aware of the fact that this assumption is sometimes naïve. Researchers have discovered, for instance, that there are often implausibly large spikes in population counts just above the threshold that makes a community eligible for a program, indicating that the population counts were manipulated to make politically favored communities eligible for the program. This undermines causal inference based on comparisons of outcomes in the two types of communities, since the program recipients might enjoy other benefits associated with the fact that they are in the good graces of the central government. In short, this type of design requires that evaluators have a very high

degree of confidence in the independence and capacity of those producing the data used to determine eligibility.

When randomization is not possible, an additional second-best strategy is to carefully match communities that have received the program with those that have not. This is an area in which statisticians have recently made significant progress (see, e.g. Diamond and Sekhon 2012). The idea of matching is to find, for every treated locality, at least one non-treated locality with identical observable characteristics, and then contrast the outcomes among treated and non-treated units.

This technique depends heavily on the availability of high-quality data on all of the possible confounders, and the presence of sufficiently similar treated and non-treated localities in the study. These can be significant hurdles, since the confounders that are most worrisome to program evaluators are often those that are least amenable to capture in a census or survey: for instance subtle differences in cultural practices, norms, or historical legacies.

6.3 Deciding Geographic Level of Interventions

At what geographic level should intervention take place? This is a crucial question in many sectors. In countries where decentralization is window dressing and power resides at the center, it is difficult to know whether interventions should focus on the central level, individual regional or local governments, or at points of service provision like schools or health clinics. Because evaluation teams place a high value on statistical power, they often have a fondness for interventions at a very low geographic level because it increases the number of potential units for analysis. Specifically, evaluation teams are attracted to rural villages, and thus, studies of villages or groups of villages are dominant in the literature.

This approach can have serious disadvantages. In many developing countries, urbanization has been proceeding rapidly, and some of the most populous and challenging places are being ignored. Moreover, villages are often not important units of governance, and may not play a very important role in budgeting, planning, or service provision. In some cases, the twin goals of experimental control and statistical power have generated the creation of new, parallel institutions of governance (e.g. general or sector-specific community meetings) without much thought to how these might engage with existing levels of government that have power to raise taxes, specify priorities, draw budgets, spend public funds, or provide services.

In the DG area, program designers are more likely to focus on building capacity, accountability, and participation at formal layers of government that play a more substantive role in determining priorities (e.g. counties, districts, or even regions). Most of the countries where USAID works have a relatively small number of such units, however, and most programs focus on fewer than half of the units. Unfortunately, this level of engagement, even if correct, considering the goals of the program designers and the needs of the country, can rob the evaluation team of the statistical power needed for a plausible study using matching or randomization.

This is a vexing problem, but one that can be resolved in innovative ways on a case-by-case basis. The solution must be found at the design phase in a spirit of collaboration that attempts to reach the frontier of the trade-off (if there is one) between a substantive evaluation and learning agenda.

6.4 What Outputs Should Be Measured?

We suggest that like decentralization, responsiveness and accountability should ultimately be viewed as means to an end. The key claim in this area is that reforms and interventions can enhance service provision by improving the structure of incentives among citizens, politicians, and bureaucrats. A program should not be deemed a success because it has created energetic and well-attended meetings, increased electoral participation, or even if it results in the removal of some corrupt officials. Instead, we propose that program goals focus on three, clearly identifiable *outcomes* of local government: First, the quality, quantity and efficiency of public services themselves; second, the distribution of access to quality public services across the citizenry; and third, citizen satisfaction with government.

Program goals and subsequent impact evaluations should focus on service delivery. It is important to focus on the quality and quantity of services, as well as the extent to which they match citizen demands, but it is important to also focus on the unit cost. If efforts at accountability promotion are successful at inducing stronger effort and less corruption among officials, this should show up in a reduced unit cost of service provision. A related concept is leakage, the amount of public funds that are lost on the way from budgetary outlay to final service provision. This can be estimated by comparing central and local budget and expenditure data, but it often requires more detailed information on the quality of public services.

Another goal that receives surprisingly little discussion and evaluation effort is equity in service provision. A project should not be deemed successful if it improves water quality only for the homes of the associates of powerful elites, or only for more affluent households living in close proximity to the center of the village.

The third goal is to promote satisfaction with governance procedures and the outputs of government. Care should be taken not to conflate citizen satisfaction with high quality, low cost service provision. In fact, there may be a wide range of settings in which interventions increase citizen satisfaction while having no impact or perhaps even decreasing the quality or efficiency of service provision, as in Olken (2010).

The quality, quantity and distribution of public services can be measured through typical citizen and administrative surveys; though it is worth noting that getting representative subnational samples will require large survey efforts. Cell phones and cameras also offer the potential to provide real-time data on service hours, wait times, stocks, etc. The efficiency of public services can be measured using performance audits and differences in pre- and post-programming services. Indeed, one important advantage of rigorous

program designs and evaluations is that they provide a clear picture of the return on aid investments that can be used to inform policymaking. Citizen satisfaction can be measured using surveys and election outcomes.

To conclude, some of the interventions described above lend themselves to straightforward empirical metrics: electoral turnout, information available to voters, attendance at meetings, completed audits, and electoral consequences of revealed corruption (or honesty). Such data are quite valuable, but they should be viewed as intermediate inputs in service of the broader goal of high quality, low-cost collective goods that are in accordance with the needs and demands of citizens as well as improved citizen satisfaction.

Conclusion

Empirical research on responsiveness and accountability in local governance and service delivery is still in its early stages. As USAID and its partners continue to make substantial investments in programming in this area, it would be desirable to come up with a set of simple, general claims about what works and what does not.

Research thus far allows no general, one-size-fits-all solutions. Interventions that seem to achieve good results in one context often fall flat in another. What works well with health care monitoring might not work with teachers. What works in a small village might not work in a town. Yet with each empirical study, the community of scholars and practitioners learns more about the conditions under which specific programs are most likely to achieve their desired results and the mechanism through which they do so. The accumulation of knowledge is slow but measurable.

USAID is in an excellent position to contribute to this growing base of knowledge. Interventions and evaluations based on clear and specific hypotheses within the context of a broad USAID framework can help USAID learn valuable lessons about the conditions under which its favored interventions are most likely to work, and at what cost.

This paper has emphasized the importance of carefully articulating a program theory linking specific interventions to specific measurable outcomes. In order to develop the requisite nuance and detail about conditions for success, it is necessary to focus on specific causal mechanisms. Drawing on the research reviewed in this paper, programs should be designed in ways that allow evaluators the opportunity to assess the causal impact of specific interventions. The interventions should also be tied to an overall USAID framework such that each project's contribution to general organizational knowledge is explicit from the beginning.

In many cases, successful evaluations will require simpler and more focused interventions than have been favored in the past. In this document we have outlined a large number of potential interventions, but this should in no way suggest that they should be combined in a haphazard way. The many potential interventions reflect the complexity of accountability relations, and one or the other intervention might be more appropriate for a given country

context. However, we reiterate that minimizing the number of interventions is key to program evaluability. Every time a variety of loosely related interventions predicated on different assumptions, goals, and causal mechanisms are bundled together and implemented uniformly in one or two favored regions without concern for the evaluation agenda, an important opportunity is lost for both USAID and the broader research community. The careful assignment of localities or individuals into treatment and control groups must be part of the initial process of program design.

More generally, program evaluation should not be viewed as an afterthought. Rather, it must be an integral part of the process of program design in order to 1) facilitate the clearest possible link between specific interventions and specific outcomes and 2) resolve specific issues like randomization, phased roll-outs, or matching before it is too late.

While this document has laid out some of the central hypotheses in the literature and provided skeletons of some potential interventions and evaluations, much of the design work is inherently context-specific. The best studies are ones that show creativity in adapting basic ideas to the local context.

At the same time, given what we are learning about the context-specificity of some results, there is no shame in replicating previous studies in new settings. By attempting similar interventions in a variety of settings, both within and across countries, USAID can add measurably to the growing base of knowledge about what works under which conditions.

References

- Achen, Christopher H., and Larry M. Bartels. 2004a. *Blind Retrospection: Electoral Responses to Drought, Flu, and Shark Attacks*. Princeton University, Manuscript.
- . 2004b. *Musical Chairs: Pocketbook Voting and the Limits of Democratic Accountability*. Princeton University, Manuscript.
- Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Julia Tobias. 2012. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." *American Economic Review* 102(4): 1206–1240.
- Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. 2009. "Report Cards: The Impact of Providing School and Child Test-scores on Educational Markets." Pomona College, World Bank, and Harvard University, Working Paper.
- Banerjee, Abhijit, Angus Deaton, and Esther Duflo. 2004. "Wealth, Health, and Health Services in Rural Rajasthan." *American Economic Review* 94(2): 326–330.
- Banerjee, Abhijit V., Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. 2010. "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India." *American Economic Journal. Economic Policy* 2(1): 1–30.
- Bardhan, Pranab K., Dilip Mookherjee, and Monica Parra Torrado. 2010. "Impact of Political Reservations in West Bengal Local Governments on Anti-Poverty Targeting." *Journal of Globalization and Development* 1(1).
- Beaman, Lori, Esther Duflo, and Petia Topalova. 2011. "Political Reservation and Substantive Representation: Evidence from Indian Village Councils." Brookings Institution Press and The National Council of Applied Economic Research.
- Berry, Christopher R. 2009. *Imperfect Union: Representation and Taxation in Multilevel Governments*. New York: Cambridge University Press.
- Besley, Timothy. 2006. *Principled Agents? The Political Economy of Good Government*. Oxford: Oxford University Press.
- Björkman, Martina, and Jakob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda." *Quarterly Journal of Economics* 124(2): 735–769.
- Chattopadhyay, Raghavendra, and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72(5): 1409–1443.

Cole, Shawn, Andrew Healy, and Eric Werker. 2012. "Do Voters Demand Responsive Governments? Evidence from Indian Disaster Relief." *Journal of Development Economics* 97(2): 167–181.

Diamond, Alexis and Jasjeet Sekhon. 2012. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." *Review of Economics and Statistics*, forthcoming.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101(5): 1739–1774.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2006. "Using Randomization in Development Economics Research: A Toolkit." Centre for Economic Policy Research, Working Paper.

Ferraz, Claudio, and Frederico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics* 123(2): 703–745.

———. 2011. "Electoral Accountability and Corruption: Evidence from the Audits of Local Governments." *American Economic Review* 101(4): 1274–1311.

Finan, Frederico, and Laura Schechter. 2012. "Vote-Buying and Reciprocity." *Econometrica* 80(2): 863–881.

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven: Yale University Press.

———. 1990. "Information and Rationality in Elections." In *Information and Democratic Processes*, eds. John Ferejohn and James Kuklinski. Urbana: University of Illinois.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review* 102(01): 33–48.

Gerber, Alan S., Gregory A. Huber, David Doherty, and Conor M. Dowling. 2012. "Is There a Secret Ballot? Ballot Secrecy Perceptions and Their Implications for Voting Behaviour." *British Journal of Political Science*.

Humphreys, Macartan, and Jeremy Weinstein. 2007. "Policing Politicians: Citizen Empowerment and Political Accountability in Uganda." Presented at the Annual Meeting of the American Political Science Association.

Keefer, Philip, and Stuti Khemani. 2012. "Do Informed Citizens Receive More...Or Pay More? The Impact of Radio on the Government Distribution of Public Health Benefits." World Bank Policy Research Working Paper No. 5952.

Khwaja, Asim Ijaz. 2004. "Is Increasing Community Participation Always a Good Thing?" *Journal of the European Economic Association* 2(2/3): 427–436.

———. 2009. "Can Good Projects Succeed in Bad Communities?" *Journal of Public Economics* 93(7–8): 899–916.

Martinez-Bravo, Monica, Gerard Padro i Miquel, Nancy Qian, and Yang Yao. 2012. "Do Local Elections in Non-Democracies Increase Accountability? Evidence from Rural China." NBER Working Paper # 16948.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119(1): 39–77.

North, Douglass Cecil, John Joseph Wallis, and Barry R. Weingast. 2009. *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. New York: Cambridge University Press.

Olken, Benjamin A. 2010. "Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia." *American Political Science Review* 104(02): 243–267.

Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115(2): 200–249.

Paler, Laura. 2012. "Keeping the Public Purse: An Experiment in Windfalls, Taxes and the Incentives to Restrain Government." Working Paper, The Center for Global Development.

Pande, Rohini. 2003. "Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India." *American Economic Review* 93(4): 1132–1151.

Persson, Torsten, and Guido Enrico Tabellini. 2000. *Political Economics: Explaining Economic Policy*. Cambridge, MA: MIT Press.

Popkin, Samuel L. 1994. *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. Chicago: University of Chicago Press.

Riker, William H. 1982. *Liberalism against Populism: A Confrontation between the Theory of Democracy and the Theory of Social Choice*. San Francisco: W.H. Freeman.

Susan Stokes. 2005. "Perverse Accountability: A Formal Model of Machine Politics with Evidence from Argentina." *American Political Science Review* 99 (3): 315-325.

Tsai, Lily L. 2007. *Accountability without Democracy: Solidary Groups and Public Goods Provision in Rural China*. New York: Cambridge University Press.

Tufte, Edward R. 1978. *Political Control of the Economy*. Princeton: Princeton University Press.

Véron, René, Glyn Williams, Stuart Corbridge, and Manoj Srivastava. 2006. "Decentralized Corruption or Corrupt Decentralization? Community Monitoring of Poverty-Alleviation Schemes in Eastern India." *World Development* 34(11): 1922–1941.

Vicente, Pedro C., and Leonard Wantchekon. 2009. "Clientelism and Vote Buying: Lessons from Field Experiments in African Elections." *Oxford Review of Economic Policy* 25(2): 292–305.

Wantchekon, Leonard. 2003. "Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin." *World Politics* 55(3): 399–422.

World Bank. 2004. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: World Bank Publications.

———. 2010. *Africa Development Indicators 2010*. Washington, DC: World Bank Publications.

Additional Resources:

Heinrich, Carolyn J. 2002. "Outcomes-Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness" *Public Administration Review* 62 (6): 712-725.