

# The ANNALS of the American Academy of Political and Social Science

<http://ann.sagepub.com>

---

## Democracy, Governance, and Randomized Development Assistance

Devra C. Moehler

*The ANNALS of the American Academy of Political and Social Science* 2010; 628; 30

DOI: 10.1177/0002716209351505

The online version of this article can be found at:  
<http://ann.sagepub.com/cgi/content/abstract/628/1/30>

---

Published by:



<http://www.sagepublications.com>

On behalf of:



[American Academy of Political and Social Science](#)

Additional services and information for *The ANNALS of the American Academy of Political and Social Science* can be found at:

**Email Alerts:** <http://ann.sagepub.com/cgi/alerts>

**Subscriptions:** <http://ann.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations** <http://ann.sagepub.com/cgi/content/refs/628/1/30>

# Democracy, Governance, and Randomized Development Assistance

By  
DEVRA C. MOEHLER

Can field experiments be productively employed to study the impact of development assistance on democracy and governance (DG) outcomes? A small but growing number of practitioners and scholars, often working in partnership, are inventing a new research domain at the intersection of evaluation and political science. The article reviews recent and ongoing DG field experiments, and it offers lessons about the prospects and obstacles to the future development of a useful body of experimental evidence on the political economy of development.

*Keywords:* democracy; governance; democratization; field experiment; randomized controlled trial; democracy promotion; impact evaluation

Since the 1980s, democracy and governance (DG) promotion programs have proliferated at an ever-increasing rate. Governments, international financial institutions, multilateral bodies, and international and domestic nongovernmental organizations provide assistance targeted at inducing democratic transitions in authoritarian polities; consolidating democracy where it exists; and increasing government effectiveness, transparency, and responsiveness to citizens across all regime types. Donors currently support explicit efforts to expand press freedom, establish rule of law, enhance government service delivery, strengthen civic culture, increase voter turnout, ensure free and fair elections, reduce corruption, and improve upon a myriad of other DG goals (Gershman

*Devra C. Moehler is an assistant professor at the Annenberg School for Communication at the University of Pennsylvania. In 2008-2009, she served as a Democracy Fellow in the Office of Democracy and Governance at the U.S. Agency for International Development, where she contributed to the development of a new pilot program in impact evaluation. Her research interests include political communication, public opinion, democratization, democracy and governance promotion, and African politics.*

DOI: 10.1177/0002716209351505

and Allen 2006). In addition, development agencies are interested in the link between their economic or social aid programs, on one hand, and DG, on the other. Unfortunately, systematic evidence on the political impact of development assistance lags behind the rapid growth of DG-related development programming.

Can field experiments be productively employed to study the impact of development assistance on DG outcomes?<sup>1</sup> Key donor agencies—such as the U.S. Agency for International Development (USAID) and the World Bank—hope that randomized controlled trials (RCTs) can provide robust evidence of impact to inform program decisions and justify resource allocations (Savedoff, Levine, and Birdsall 2006). Academics recognize that field experiments offer powerful tools for testing causal arguments about the drivers of political change in developing countries (Humphreys and Weinstein 2009). A small but growing number of practitioners and scholars, often working in partnership, are inventing a new research domain at the intersection of evaluation and political science. This article reviews recent and ongoing DG field experiments and offers some lessons about the prospects and obstacles to the future development of a useful body of experimental evidence on the political economy of development.

This article focuses attention on randomized field experiments in the developing world that explicitly address (1) DG outcomes of social or economic development assistance, (2) social and economic outcomes of DG promotion programs, and (3) DG outcomes of DG programs. In this article, all field experiments of development assistance with DG intervention and/or outcomes are referred to by the shorthand title “DG field experiments.” This article focuses on research that is at the intersection of three larger bodies of work. First, it reviews experiments carried out by development agencies and scholars of political economy of development. This research is certainly informed by experiments in advanced industrial democracies, but it concentrates on interventions in poor countries and in polities at the lower end of the democracy or governance spectrum. Second, it addresses a new subfield of experimental studies of development aid, in which research on education, health, and economic interventions and outcomes has heretofore dominated research on DG. Third, it concentrates on a portion of the more general category of impact evaluations of DG outcomes. Impact evaluations consider the counterfactual and include experimental designs, such as the ones considered here, as well as quasi-experimental approaches.<sup>2</sup> While this inquiry is focused on the more limited domain of field experiments of DG interventions and outcomes in the developing world, many of the observations and lessons are relevant to the broader fields of inquiry mentioned here.

The first section of this article describes current institutional support for DG field experiments. The second section provides an overview of current studies. The third section offers some warnings and lessons about challenges to successful implementation of DG field experiments, and the final section concludes by looking to the future.

## Support of Democracy and Governance Field Experiments

To date, much of the institutional support for DG field experiments has come from academic institutions. Most authors of DG field experiments are members of political science or economics departments at research universities.<sup>3</sup> Notably, most academics involved in this enterprise work collaboratively with practitioners. Academics have partnered with large donors (such as USAID and the World Bank), international NGOs (such as International Rescue Committee, ActionAid, and the Carter Center), or local civil society organizations (such as Africa Leadership Institute in Uganda).

The World Bank has several programs to support field experiments, some of which explicitly consider the effect of development assistance programs on DG outcomes, such as the Africa Impact Evaluation Initiative, the Development Research Group, the Social Development Department, the Spanish Trust Fund for Impact Evaluation, and the Development Impact Evaluation Initiative. World Bank projects are primarily targeted at economic outcomes. Nonetheless, growing numbers of experimental impact evaluations are being designed to test the effects on DG outcomes of activities such as citizen cooperation and political participation, corruption, government accountability, and public service delivery.

USAID has a new pilot impact evaluation program in the Office of Democracy and Governance. The pilot program follows the recommendations of a National Academies of Science committee (Committee on the Evaluation of USAID Democracy Assistance Programs 2007). To my knowledge, this is the only donor program that is exclusively targeted at impact evaluations of DG-assistance programs. While still in its infancy, this initiative aims to increase the number of USAID-supported DG programs that are subject to experimental (and quasi-experimental) impact evaluation. It has the potential to contribute significantly to the body of work in this new field.

## Democracy and Governance Experiments Today

What aspects of DG interventions and outcomes are currently being examined with field experiments? To answer this question, I compiled a list of 41 studies that have been completed or are currently under way (see Table 1). This list was constructed from published work, conference or other public presentations, and the Web sites of large development donors or organizations<sup>4</sup> that promote impact evaluations in the development field.<sup>5</sup>

Just over half of the field experiments took place in Africa (22 out of 41). Nine were conducted in India; three in Indonesia; and the remainder in Mexico, Brazil, China, Georgia, and Afghanistan. The regional concentrations reflect the expertise of academic pioneers in these areas and the investments of the Africa

TABLE 1  
Democracy and Governance Field Experiments

|    | Authors  | Country             | Topic                  |
|----|--|---------------------|------------------------|
| 1  | Arcand   | Gambia              | CDD                    |
| 2  | Arcand   | Angola              | CDD                    |
| 3  | Banerjee, Banerji, Duflo, Glennerster, & Khemani                                     | India               | Government performance |
| 4  | Banerjee, Chattopadhyay, Duflo, & Keniston   | India               | Government performance |
| 5  | Banerjee, Green, Green, & Pande  | India               | Elections              |
| 6  | Beaman, Chattopadhyay, Duflo, Pande, & Topalova                                      | India               | Other: Quota           |
| 7  | Beath, Christia, & Enikolopov  | Afghanistan         | CDD                    |
| 8  | Bertrand, Djankov, Hanna, & Mullainathan   | India               | Government performance |
| 9  | Björkman & Svensson  | Uganda              | Government performance |
| 10 | Chattopadhyay & Duflo  | India               | Other: Quota           |
| 11 | Collier & Vicente  | Nigeria             | Elections              |
| 12 | De La O Torres   | Mexico              | Elections              |
| 13 | Driscoll & Hidalgo   | Georgia             | Elections              |
| 14 | Duflo & Hanna  | India               | Government performance |
| 15 | Duflo, Fischer, & Chattopadhyay  | India               | Other: Quota           |
| 16 | Fearon, Humphreys, & Weinstein   | Liberia             | CDD                    |
| 17 | Ferraz & Finan   | Brazil              | Elections              |
| 18 | Glennerster, Miguel, & Zhou  | Sierra Leone        | CDD                    |
| 19 | Goldstein  | Indonesia           | Government performance |
| 20 | Guan & Green   | China               | Elections              |
| 21 | Gugerty & Kremer   | Kenya               | Other                  |
| 22 | Humphreys  | DRC                 | CDD                    |
| 23 | Humphreys & Weinstein  | Uganda              | Elections              |
| 24 | Humphreys, Masters, & Sandu  | São Tomé & Príncipe | Other                  |
| 25 | Hyde   | Indonesia           | Elections              |
| 26 | Ichino & Schündeln   | Ghana               | Elections              |
| 27 | King, Gakidou, Ravishankar, Moore, Lakin, Vargas, Tellez-Rojo, Avila, Avila, & Llama | Mexico              | Other                  |
| 28 | Kremer, Moulin, & Namunyu  | Kenya               | Government performance |
| 29 | Lierl  | Guinea              | CDD                    |
| 30 | Miguel, Kremer, & Thornton   | Kenya               | Other                  |
| 31 | Moehler  | Uganda              | Government performance |
| 32 | Olken  | Indonesia           | CDD                    |
| 33 | Olken  | Indonesia           | CDD                    |
| 34 | Özler, Baird, & McIntosh   | Tanzania            | CDD                    |
| 35 | Paluck   | Sudan               | Other                  |
| 36 | Paluck   | DRC                 | Other                  |
| 37 | Paluck   | Rwanda              | Other                  |
| 38 | Pandey, Goyal, & Sundararaman  | India               | Government performance |
| 39 | Vicente  | São Tomé & Príncipe | Elections              |
| 40 | Wantchekon   | Benin               | Elections              |
| 41 | Wantchekon   | Benin               | Elections              |

Impact Evaluation Initiative. A quarter of the field experiments took place in post-conflict settings or in settings subject to ongoing conflict, such as Eastern DRC and Afghanistan. While field experiments may be difficult to conduct in unstable environments, this has proven to be a surmountable obstacle.

In addition to regional concentrations, there are also substantive topics that dominate the sample. The studies could be classified in several ways based on the motivating development hypothesis, type of intervention, or outcome of interest. Rather than duplicate existing reviews that focus on theories (Humphreys and Weinstein 2009), this article groups studies according to interventions, a framework more consistent with the approach of development practitioners. Among the 41 field experiments, there are three main types of interventions: elections, community-driven development (CDD), and reforms to improve public service delivery.

### *Elections*

Twelve of the 41 DG field experiments focus on electoral interventions or outcomes. The most common types of electoral interventions are randomized local-level voter education or mobilization campaigns held prior to national elections. The experimental work on elections provides evidence of the ways in which information messages change citizen electoral behavior and by extension electoral outcomes. Most of the studies measure citizen behavior through surveys and election data, while a few also invent behavioral tests (Collier and Vicente 2008; Driscoll and Hidalgo 2009). Some studies also estimate the effects of interventions on the behavior of leaders (e.g., Humphreys and Weinstein 2007; Hyde 2008).

Several experiments evaluate campaigns designed to deter or punish electoral malfeasance. Vicente (2007) found that a randomized leaflet door-to-door campaign against vote buying, sponsored by the São Tomé and Príncipe Electoral Commission, increased negative perceptions of vote buying, but also reduced turnout in targeted census areas. Collier and Vicente (2008) investigate a campaign against political violence that was randomized across neighborhoods and villages of six Nigerian states.<sup>6</sup> They establish that the campaign reduced electoral violence as recorded by perception surveys and independent data sources and, as a result, increased citizen turnout and reduced votes for the opposition leaders, the main perpetrators of election violence.<sup>7</sup> Driscoll and Hidalgo (2009), in cooperation with the National Democratic Institute, target randomly assigned precincts with instruction for voters to either check public voter rolls or make complaints about problems observed on Election Day. They measure the impact of the canvassing and pamphleting information using electoral data, survey panel data, and NGO reports. Hyde (2008) randomly assigned international observers to locations for Election Day monitoring in Indonesia's 2004 presidential election to study their effect on fraud and malfeasance at the precinct level. Ichino and Schündeln (2009) randomly assign Ghanaian domestic observers to constituencies during the July and August 2008 voter registration period to estimate the effect of monitoring on the percentage change in the number of registered voters.

Other experiments evaluate interventions designed to encourage more active or civic-oriented citizen participation in elections. Guan and Green (2006) found that a non-coercive mobilization campaign had a positive effect on voter turnout in China's 2003 elections for the county- and district-level People's Congress; the door-to-door get out the vote campaign targeted a random selection of students in a Peking University precinct. Banerjee, Green, et al. (n.d.) randomly assigned local units in India to receive campaigns designed to (1) reduce caste-based voting, (2) persuade citizens to vote against corrupt politicians, or (3) mobilize women to vote.<sup>8</sup> Wantchekon (2003, 2008) conducted two field experiments where political parties altered their campaign messages in a random selection of villages to convey broad public goods electoral platforms as opposed to clientelist platforms. He found that clientelist messages were more effective at winning votes in the 2001 experiment, but the clientelist advantage disappeared in 2006 when refined public goods proposals were delivered in town meetings. Humphreys and Weinstein (2007) test whether informing constituents about the performance of a random selection of members of the Ugandan Parliament induces performance-based voting and greater MP effort.

Finally, a few scholars studied the possible unintended electoral effects of randomized policy programs; De La O Torres (2008) ascertained that a randomly assigned Education, Health, and Nutrition Program of Mexico (*Progresa*) increased turnout and the incumbent vote share. Ferraz and Finan (2007) made use of randomly assigned municipal expenditure audits in Brazil to establish that public release of information about municipal governments helped voters to punish corrupt leaders.

While firm conclusions await additional evidence (including that of the experiments still in progress), on balance the existing studies suggest that even limited interventions prior to elections can significantly influence the electoral behavior of citizens and leaders. The results are heartening to those who seek to improve the quality of elections. They also call into question simple voting models based solely on ethnic or economic characteristics; new information and appeals to the common good can induce citizens to alter their electoral behavior.

### *Community-driven development*

Ten of the 41 field experiments evaluate the effects of CDD programs, which typically seek to foster participatory processes for selecting, allocating, and monitoring local public goods projects. The prevalence of field experiments of CDD programs results from (1) the ease of randomizing treatments, since most CDD programs are implemented at the village level; (2) the growth of CDD programs as key components of donor-funded poverty reduction efforts; and (3) the World Bank's focus on CDDs for one of several clusters of impact evaluation. Most of the experiments evaluate the impact of the CDD programs on public goods provision, citizen welfare, and poverty. However, many also investigate DG outcomes such as social cohesion and collective action, positive attitudes toward government, civic culture, public participation, and avoiding elite capture (e.g.,

Arcand n.d.-a, n.d.-b; Beath, Christia, and Enikolopov 2008; Fearon, Humphreys, and Weinstein 2009; Casey, Glennerster, and Miguel 2009; Humphreys n.d.; Lierl 2009; Özler, Baird, and McIntosh n.d.). These studies test whether external assistance can facilitate the development of democratic attitudes, practices, and institutions at the local level.

For example, Fearon, Humphreys, and Weinstein (2009) evaluate the effect of an International Rescue Committee–led community-driven reconstruction project in post-conflict Liberia. The treatment, randomized across 42 of 83 deserving communities (usually small clusters of villages), introduced democratic, community-level institutions for making and implementing decisions about local public goods. Anonymous public goods games played in treatment and control villages indicate that the treatment increased collective action capacity. In a similar experiment, Lierl (2009) randomly assigned one rural development community from each of 52 pairs to receive the program in 2009 and the other in 2010. The study, still in progress, will estimate the impact of participatory processes on citizen–authority relations at the local level, including trust in authorities, capacity for local collective action, and the quality of interaction between citizens and local authorities.

In addition to testing the effects of receiving or not receiving CDD-type program funds, several experiments randomize election rules or decision-making procedures. These variations are designed to test the influence of different institutional arrangements on citizen satisfaction, interest aggregation and representation, and agent accountability.

Beath, Christia, and Enikolopov (2008) query the effect of single-member versus multi-member constituencies; Humphreys (n.d.) analyzes the effect of introducing electoral quotas; and Olken (2007b) and Beath, Christia, and Enikolopov (2008) compare secret-ballot elections with deliberative public meetings as processes for selecting projects. Olken (2007a) compares the effect of top-down versus bottom-up monitoring on corruption in 608 Indonesian village road projects. He finds that dramatically increasing the frequency of top-down government audits reduced the discrepancy between official project costs and independent engineers' estimates of costs, but increasing grassroots participation in monitoring had little average impact.

Most of the CDD-type field experiments are still in progress. Initial indications are that the programs' political effects are at least as great as, if not greater than, their economic effects. In addition, the experiments promise important insights about best practices in local-level institutional design.

### *Public service delivery and government performance*

Nine of the 41 field experiments focus on interventions designed to improve government performance and service delivery. Several interventions seek to increase the beneficiaries' participation in monitoring public services and government officials. Banerjee et al. (2006) show that interventions to help village education committees monitor education service providers in India were not

effective, though Pandey, Goyal, and Sundararaman (2008) find that a different community-based information campaign in India did improve school performance, at least in the short term. Moehler (2008) broadcasts results of local government audits over the radio for a random selection of subcounties in Uganda and notes changes in the knowledge, attitudes, and behaviors of citizens and in the performance of public officials. The design also allows for a comparison of the efficacy of information bulletins versus talk-show formats. Banerjee, Chattopadhyay, et al. (n.d.) consider the effects of greater community interaction with police (and other reforms) on police performance. Björkman and Svensson (2007) document significant gains in health service provision when citizens monitored healthcare providers with community report cards. Average clinic utilization was greater, infant weight was higher, and the number of deaths among children under five was markedly lower in the catchment areas of 25 randomly targeted health clinics than in the catchment areas of control clinics. The evidence suggests that the treatment communities became more extensively engaged in monitoring the health unit and that healthcare providers exerted more effort. Like many of the CDD and election experiments, these studies test the hypothesis that greater citizen access to information about officials increases accountability and as a result improves government performance. To date the evidence on the hypothesis is mixed. Additional studies should help to define the conditions under which the hypothesis does and does not receive support.

Others employ field experimental methods to investigate alternative reforms to improve public provision of services. Kremer, Moulin, and Namunyu (2003) draw lessons about the perverse incentives of partial decentralization in Kenya's education system based on experimental evidence of budget-neutral reforms and student transfers. Bertrand et al. (2006) gain insights into the causes of corruption through their experiment in obtaining driver's licenses in India. Duflo and Hanna (2005) show that performance-based pay reduces teacher absenteeism in India. Goldstein (2009) tests the effects of community-based legal education and aid on enforcement of rights, dispute resolution, and welfare in Indonesia. Community legal aid posts are randomized across subdistricts and paralegals are randomized across villages. Outcomes of interest in this study include beneficiaries' ability to claim rights, enforce rights, and resolve disputes through nonviolent means as well as reduced violence and conflict, increased safety and security, and improved household welfare.

### *Other democracy and governance field experiments*

The remaining ten experimental studies address a variety of other themes. Three exploit a randomized policy experiment in India, which demonstrates significant effects of gender and minority quotas in village councils (Beaman et al. 2008; Chattopadhyay and Duflo 2004; Duflo, Fischer, and Chattopadhyay 2005). Gugerty and Kremer (2008) document the detrimental effect on the disadvantaged from outside funding for women's community associations in Kenya. A randomized delayed-treatment design was used to compare 40 treated women's

associations with 40 that had not yet received assistance. While there were no significant differences between treatment and control groups in the desired outcomes (organizational strength, participation, assistance to neighbors, or contribution to local public goods), outside funding changed the nature of membership and leadership in the community associations, further marginalizing the relatively disadvantaged, older, uneducated women. Humphreys, Masters, and Sandu (2007) find that the outcomes of deliberative forums in São Tomé were systematically affected by the preferences and demographic characteristics of group leaders. King et al. (2007) provide an example of how randomization of a public policy intervention can facilitate evaluation even in the face of political manipulation. Paluck (2008, 2009, n.d.) employs experimental methods in Rwanda, the DRC, and South Sudan to study the effects of educational radio soap operas and listening group discussions on reconciliation, tolerance, community cohesion, and civic culture. Finally, Miguel, Kremer, and Thornton (2008) make use of an experimental intervention that provided scholarships to some students but not others to study the political effects of education. They find that education leads not only to greater political knowledge but also to less satisfaction with Kenya's democracy, less respect for authority, less optimism about the quality of government, and greater tolerance for violence.

This brief canvass of DG field experiments illustrates the breadth of work that currently exists and suggests avenues for future inquiry. Although still in its infancy, the enterprise of DG field experiments could potentially generate novel evidence to address important questions of interest to academics as well as to donors and implementing partners. In addition to evaluating interventions in the realms of elections, community collective action, and reforms to improve the quality and quantity of public services, other common DG interventions are also amenable to field experiments. Interventions to improve citizen access to impartial justice, journalistic standards of reporting, civil society organizing, capacity-building in local governments, empowerment of civil society groups, and training of officials in party branch offices are amenable to impact evaluation. The resulting evidence could help resolve critical theoretical debates about the influence of rule of law, independent press, civil society, decentralization, and political parties.

## Democracy and Governance Field Experiment Challenges

Field experiments generate high internal validity because they address the problem of selection bias with random assignment of target and control units. However, the particular challenges of executing field experiments lead researchers to focus on certain types of interventions and limit the types of evidence collected through the field experiment methodology. While a more narrow focus is not an acute problem for science or program evaluation, it does constrain the lessons one can accurately learn from RCTs at this time.

This section provides some preliminary thoughts about key challenges to the implementation of DG field experiments based on the author's work during the first eight months of the pilot impact evaluation project in the Office of Democracy and Governance at USAID. These challenges are not unique to DG experiments; they affect field experiments more generally to varying degrees. While familiar to experimentalists, they may not be as obvious to political economy of development scholars and practitioners. Recognizing the current limitations to the evidence produced by field experiments can help scholars and practitioners identify appropriate domains for generalizing theoretical and programmatic findings. It is hoped that shedding light on these challenges will spur scholars and practitioners to find ways to overcome them.

### *Level of intervention and analysis*

The level of intervention represents the biggest potential bias of impact evaluations. The precision of estimates depends on the treatment of many units, and the need for large numbers of units in experimental research pushes researchers to study low-level interventions.<sup>9</sup> The vast majority of the 41 studies under review evaluate interventions at the level of a village or neighborhood (or similarly at a local school, clinic, or police station). These studies offer important lessons for other community-level interventions to improve welfare—the site of a significant amount of current development practice. However, many DG interventions are targeted at higher levels of government. Furthermore, our theories of political economy of development are typically based on higher-order institutions and interactions. Community-based interventions are not necessarily a problem for theory development if they are capable of testing the micro-foundations of theories. The key problem is that the unique contextual features of a village or neighborhood could interfere with our ability to generalize from micro-level results. In villages or neighborhoods, citizens often know each other, share strong informal norms, can observe outcomes, and have reason to expect repeated interactions. Voting rules or mechanisms for decision-making that work well among acquaintances may not yield optimal results beyond such intimate settings. Lessons about citizen participation and monitoring at the point of service delivery may not translate into best practices for building accountability at higher levels of government. The context of the interventions needs to be considered when scholars or practitioners draw inferences from the evidence.

While clearly revealing a tendency toward village-level interventions, the field experiments reviewed here also offer some indications that impact evaluations can potentially address higher-level interventions and theories. Humphreys and Weinstein (2009) study the effects of public exposure on MP behavior; Moehler's (2008) experiment uses radio broadcasts about subcounty governments;<sup>10</sup> Ferraz and Finan (2007) show effects of municipal audits. Researchers should continue to push toward higher-level interventions that might be more analogous to national-level processes.

### *Complexity of intervention*

Another challenge results from the complexity of the DG programs, which tend to combine many different activities targeted at a diverse array of beneficiaries. There are good reasons why programs tend to include a bit of everything thought to affect target objectives; accumulated experience suggests that a holistic approach to DG programming is beneficial, and sufficient evidence does not yet exist to allow practitioners to intelligently select from the range of possible actions. Regardless of the reasons, this multifaceted approach to programming poses significant interrelated challenges to hypothesis testing. First, complex interventions tend to occur at multiple levels or with a variety of different target populations, making it difficult to decide on an appropriate frame for random assignment. At higher levels, the numbers of units are often too small to yield sufficiently precise estimates, and at lower levels, random assignment cannot capture the effects of higher-level interventions. Second, measuring outcomes is more difficult with complex interventions due to the diversity of targets. Third, if the field experiment is designed to measure the impact of the bundle of activities, then it can be hard to determine whether or not specific activities had an impact. Such an evaluation may not offer precise evidence about cause and effect that is important for theory building and policymaking.

An enticing alternative is to test the effect of individual activities separately (and in differing combinations where possible). Unfortunately, when programs combine many different kinds of activities, the resources expended on each individual activity tend to be small, and the expected effect of each activity may be too small to measure. The limited number of treatment units involved in the typical DG intervention often prohibits separate tests of individual activities and various combinations of activities within the same program. Attempts to simplify programs to facilitate evaluation risk generating a body of evidence about atypical programs. The same is true if researchers gravitate toward the study of simple programs.

### *Synergies and coordination with other programs*

The development field is often crowded with other programs. Many times donors coordinate efforts to gain from synergies in programming or to avoid duplication of efforts; both can complicate researchers' ability to discern cause and effect. If other actors are likely to devote added attention to target units in order to benefit from synergies, then experimental results will tend to overestimate the treatment effect. If other actors avoid project areas to prevent duplication of efforts, then the experimental evidence will tend to underestimate the treatment effect. The researcher typically has no ability to influence the program decisions of other development actors, and there could be significant ethical problems in attempting to do so. Researchers thus tend to focus on the least crowded fields, or to select projects where the effects of the program under investigation are expected to dwarf the effects of other programs. These individual

solutions yield a body of evidence from field experiments in atypical programs or in unusual sites (perhaps ones where the intervention is least likely to be successful) that could also limit the external validity of findings.

### *Best performers volunteer for impact evaluation*

Implementing agencies may be more likely to volunteer their projects for impact evaluation or agree to researcher requests when they expect their programs to be effective.<sup>11</sup> Researchers also tend toward evaluating projects where they expect larger effects so as to avoid inconclusive results. Finally, the costs of rigorous impact evaluation are often easier to justify for larger and more expensive programs. These tendencies could lead to an accumulation of evidence in certain types of locations or programs, which is a problem for policymakers who try to assess the average efficacy of program type based on a biased sample of the most effective programs. It is also a problem for scholars who do not pay attention to the characteristics of their sample when drawing inferences. A careful mapping of the conditions under which all similar programs are implemented versus those under experimental investigation would help both academics and practitioners understand the limitations of valid inference.

### *Research planning under ambiguity*

Program objectives, activities, and underlying development hypotheses typically evolve over time and are only well specified once the implementation of the program has begun (and sometimes not until after that). Planning and solicitation documents specify project goals and sometimes a series of illustrative activities to achieve those goals. Donors often encourage some flexibility so that the subject expertise of implementing partners can inform the program design. After an award is made, additional changes and refinements to the program are often allowed in response to early data collection and analysis. While flexibility can make for better programming, it poses considerable difficulties for the evaluator. It is difficult (and often impossible) to define an experiment when the objectives, activities, and development hypotheses linking the two are not well specified. It can be equally problematic to wait to design the study until after the program parameters are fully clarified; the intense pressure to commence implementation as soon as possible leaves little time for adjustments and baseline data collection, and agreement can be harder to secure after awards are made. The result can be a disjuncture between the optimal research design and baseline data collected, on one hand, and the eventual contours of the intervention under investigation, on the other. Measured impact may not reflect real impact due to suboptimal research designs given the actual interventions undertaken.

Similarly, many development programs fail not because the underlying development hypothesis was faulty but because of problems with implementation. From a researcher's perspective, it can be difficult to determine with great certainty the exact nature of the treatment. Collection of data on inputs, outputs and

proximate outcomes, site visits, and qualitative process tracing can reveal clues about the actual nature of the treatment, but only imperfectly.

This focus on challenges should not obscure the immense opportunities at hand. To understand something universal and generalizable about a particular phenomenon, experiments, like most other kinds of research, need to be replicated in multiple contexts. On this account, field experiments may do better than other research approaches due to the partnerships that develop between academics and practitioners. Academics have incentives to innovate research designs, while practitioners often have incentives to replicate existing evaluation design. A beneficial division of labor may develop for field experiments of development interventions that yield more replication and allow greater external validity than is typically the case for research agendas driven solely by academics. If social scientists can find creative ways to push back the implementation obstacles discussed above, then the prospects for external as well as internal validity are good.

## The Future

Many DG interventions are not amenable to experimental methodology, and exclusive reliance on a single methodology is a recipe for failure in the academy as well as for development agencies. However, recognition of the current limitations of field experiments does not necessarily forecast the sustainability and future vitality of experimental methods in the DG development field. A body of comparative experimental evidence has begun to emerge in the areas of election campaigns, CDD programs, and reforms to improve public services, but scholars and practitioners are a long way from exhausting research possibilities in these areas. Furthermore, many theoretical questions about DG programs that are amenable to experimentation have yet to be addressed. Even as the increased use of field experiments in DG provides answers to fill current gaps in evidence, new conditions, programs, and research questions are likely to emerge that will motivate new research agendas that include field experiments.

The enterprise of DG field experiments will be constrained more by mundane challenges to successful research design and implementation than by the inherent limitations of field experiments. Field experiments will be judged more on the quality and relevance of the evidence they produce than by issues or programs they do not try to address. The author hopes that this discussion of mundane challenges will help analysts consider and appropriately address potential biases when drawing conclusions from experimental evidence.

More important, experimentalists can find ways to move beyond general tendencies that limit the usefulness of the evidence. One fruitful avenue for future research would be to push above the village, neighborhood, or point of service delivery and focus on interventions with higher levels of local government and administrative units. Many current DG interventions to deepen decentralization or improve service delivery are targeted at county, state, and provincial levels of government. Middle levels of government offer settings that are more analogous

to national governments; officials are geographically and socially distant from citizens and preside over diverse populations and complex structures. Experimental investigations of local government programs (or interventions at subnational units with line ministries) can offer better tests of theories of representative government than interventions at the community level. They are also more representative of current development practice than village or neighborhood interventions.

Additional insights can be gained by exploiting national-level changes that require facilitation to ensure implementation at the local level. For example, legal reforms in Rwanda shifted the formal set of rules from a civil law system to a more common law–based system. However, sitting judges have had little training in the use of precedent and other common law practices and thus continue to judge cases as they had under the former civil law system. Random assignment of judges and other legal professionals to training courses or technical assistance would provide insights into the effects of common versus civil law practices at the court level.

Even long-standing rules or policies may become effective in practice only in the presence of outside assistance. Random assignment of assistance to register private property, increase collection of taxes, reduce corruption, introduce incentive-based pay for civil servants, speed processing of court cases, and train police in human rights can provide unique evidence on core political science questions and current development practices. As such, even local-level interventions can be useful when analyzing national-level phenomena. With additional imagination and attention by scholars and practitioners, the current constraints on DG field experiments are likely to become new opportunities.

## Notes

1. This article uses the following terms interdependently: field experiment, randomized controlled trial, experimental impact evaluation, and randomized impact evaluation.

2. This article does not address quasi-experimental techniques such as regression discontinuity, interrupted time series, difference-in-differences, propensity score matching, pipeline, or instrumental variables; nor does it address laboratory, survey, or Internet-based experiments. It includes public goods games only if they are used to measure the effects of a randomized field intervention such as in Fearon, Humphreys, and Weinstein (2009).

3. The number of institutions employing professors or training students involved with DG field experiments is still rather small. The Abdul Latif Jameel Poverty Action Lab (PAL) has generated the largest number of DG field experiments, although evaluations of DG interventions or outcomes still make up a small proportion of PAL's total portfolio of field experiments.

4. These organizations include the Poverty Action Lab (MIT), Network of Networks on Impact Evaluation (NONIE), International Initiative for Impact Evaluation (3ie), and Innovations for Poverty Action (IPA).

5. The set of studies is undoubtedly incomplete; it is more likely to include studies by professors in political science or economics departments located in the United States or by U.S.-based organizations. In addition, there are probably some studies of education and health interventions that have lessons for public administration, civil service incentives, and service delivery that are not contained in the list.

6. The campaign sought to empower citizens to counteract local violence. ActionAid, an international NGO, called upon citizens to “vote against violent politicians” in town meetings and popular theaters and distributed campaign material.

7. They also assess the network effects of the anti-violence intervention, including the reinforcement effects on treated subjects when others in the subjects' network are also treated, and the diffusion effects from treated to untreated subjects.

8. In the anti-corruption and anti-caste voting treatments, the minimum of village or electoral precinct was randomly assigned to the treatment of puppet shows and rallies. In the women's mobilization experiment, the head village of a *gram panchayat* (collection of proximate villages) was randomly assigned the treatment of small-group meetings.

9. Sometimes one also has the option of expanding the scope of the treatment. Interestingly, the important ethical concern about denying treatment to certain populations in order to maintain a control group has not been a major issue for the pilot project to date. Instead, the opposite is more common. Resource limitations and program design lead to a small fraction of potential recipients receiving assistance; from an evaluation perspective, increasing the numbers receiving treatment rather than denial of treatment is usually desirable.

10. Subcounty population totals usually exceed 10,000 residents.

11. Despite the fact that there are not many formal rewards for subjecting programs to rigorous scrutiny, many USAID officials and implementing partners have expressed interest in the pilot impact evaluation program. It seems that the benefits of learning the effects of current programs and the opportunities for improving impact in the future outweigh the costs. It is possible that resistance will become a problem as the program expands, but it has not been a major limiting factor so far.

## References

- Arcand, Jean-Louis. n.d.-a. The FAS III Baseline Survey: Preliminary findings, Angola. Unpublished.
- Arcand, Jean-Louis. n.d.-b. Technical proposal for the Baseline Survey for the Impact Evaluation of the Gambia CDDP. Unpublished.
- Banerjee, Abhijit, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. 2006. Can information campaigns spark local participation and improve outcomes? A study of primary education in Uttar Pradesh, India. SSRN Working Paper Series.
- Banerjee, Abhijit, Raghavendra Chattopadhyay, Esther Duflo, and Daniel Keniston. n.d. Police performance, community policing, and public perception in Rajasthan, India. Poverty Action Lab Projects.
- Banerjee, Abhijit, Jennifer Green, Donald Green, and Rohini Pande. n.d. Campaigns to influence voting behavior in India. Poverty Action Lab Projects.
- Beaman, Lori, Raghavendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2008. Powerful women: Does exposure reduce bias? Paper presented at Northeastern Universities Development Consortium Conference, Boston University, Boston, MA, November 8-9.
- Beath, Andrew, Fotini Christia, and Ruben Enikolopov. 2008. *Randomized impact evaluation of Afghanistan's National Solidarity Program (NSP): Hypotheses and methodology*. <http://www.nsp-ie.org/reports/HM.pdf>
- Bertrand, Marianne, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan. 2006. Obtaining a driver's license in India: An experimental approach to studying corruption. *Quarterly Journal of Economics* 122 (4): 1639-76.
- Björkman, Martina, and Jakob Svensson. 2007. Power to the people: Evidence from a randomized field experiment of a community-based monitoring project in Uganda. World Bank Policy Research Working Paper.
- Casey, Katherine Whiteside, Rachel Glennerster, and Edward Miguel. 2009. Does community driven development improve collective action in post-war Sierra Leone? Poverty Action Lab Projects.
- Chattopadhyay, Raghavendra, and Esther Duflo. 2004. Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica* 72 (5): 1409-43.
- Collier, Paul, and Pedro C. Vicente. 2008. Votes and violence: Evidence from a field experiment in Nigeria. Households in Conflict Network (HiCN) Working Papers Series.
- Committee on the Evaluation of USAID Democracy Assistance Programs. 2007. *Improving democracy assistance: Building knowledge through evaluations and research*. Washington, DC: National Academies Press.

- De La O Torres, Ana Lorena. 2008. Do poverty relief funds affect electoral behavior? Evidence from a randomized experiment in Mexico. Paper presented at Field Experiments in Comparative Politics and Policy, University of Manchester, UK, July 1-2.
- Driscoll, Jesse, and Diego Hidalgo. 2009. Against the machine: Experimental results from Georgia's 2008 parliamentary election. Paper read at annual meeting of the American Political Science Association, Toronto, Canada, August.
- Duflo, Esther, Greg Fischer, and Raghavendra Chattopadhyay. 2005. Efficiency and rent seeking in local governments: Evidence from randomized policy experiments in India. MIT Working Papers.
- Duflo, Esther, and Rema Hanna. 2005. Monitoring works: Getting teachers to come to school. National Bureau of Economic Research Working Paper Series no. 11880.
- Fearon, James, Macartan Humphreys, and Jeremy M. Weinstein. 2009. Can outsiders build cohesion after civil war: Evidence from a field experiment in post-conflict Liberia. *American Economic Review* 99 (2): 287-91.
- Ferraz, Claudio, and Frederico Finan. 2007. Exposing corrupt politicians: The effects of Brazil's publicly released audits on electoral outcomes. Institute for the Study of Labor (IZA) Discussion Papers.
- Gershman, Carl, and Michael Allen. 2006. New threats to freedom: The assault on democracy assistance. *Journal of Democracy* 17 (2): 36-51.
- Goldstein, Markus. 2009. The mediation and community legal empowerment of the Support for Poor and Disadvantaged Areas Project in Indonesia (Spada/Mcle) impact evaluation. Unpublished work in progress. [http://siteresources.worldbank.org/INTISPMA/Resources/Training-Events-and-Materials/449365-1208956735213/Day2\\_Session3\\_MGoldstein\\_MCLE\\_IEdesign.ppt#256,1](http://siteresources.worldbank.org/INTISPMA/Resources/Training-Events-and-Materials/449365-1208956735213/Day2_Session3_MGoldstein_MCLE_IEdesign.ppt#256,1), (accessed March 21, 2009).
- Guan, Mei, and Donald P. Green. 2006. Noncoercive mobilization in state-controlled elections: An experimental study in Beijing. *Comparative Political Studies* 29 (10): 1175-93.
- Gugerty, Mary Kay, and Michael Kremer. 2008. Outside funding and the dynamics of participation in community associations. *American Journal of Political Science* 52 (3): 585-602.
- Humphreys, Macartan. n.d. Political and economic impacts of development aid in Eastern Congo. Unpublished.
- Humphreys, Macartan, William Masters, and Martin Sandu. 2007. The role of leaders in democratic deliberations: Results from a field experiment in São Tomé and Príncipe. *World Politics* 58:583-622.
- Humphreys, Macartan, and Jeremy M. Weinstein. 2007. Policing politicians: Citizen empowerment and political accountability in Africa. Paper read at annual meeting of the American Political Science Association, Chicago, IL, August.
- Humphreys, Macartan, and Jeremy M. Weinstein. 2009. Field experiments and the political economy of development. *Annual Review of Political Science* 12:367-76.
- Hyde, Susan D. 2008. Randomizing international election observation: The 2004 presidential elections in Indonesia. Paper presented at presentation at Ohio State University, Columbus, OH, May 16.
- Ichino, Nahomi, and Matthias Schündel. 2009. Political participation and electoral malpractices in Ghana: The role of domestic election observers and information. Paper presented at annual meeting of the American Political Science Association, Toronto, Canada, August.
- King, Gary, Emmanuela Gakidou, Nirmala Ravishankar, Ryan T. Moore, Jason Lakin, Manett Vargas, Martha Maria Tellez-Rojo, Juan Eugenio Hernandez Avila, Mauricio Hernandez Avila, and Hector Hernandez Llama. 2007. A politically robust experimental design for public policy evaluation, with application to the Mexican universal health insurance program. *Journal of Policy Analysis and Management* 26 (3): 479-506.
- Kremer, Michael, Sylvie Moulin, and Robert Namunyu. 2003. Decentralization: A cautionary tale. Poverty Action Lab, Paper no. 10.
- Lierl, Malte. 2009. Personal communication about the Village Community Support Program in Guinea.
- Miguel, Edward, Michael Kremer, and Rebecca Thornton. 2008. Does education change political attitudes? Evidence from a Kenyan school experiment. Paper presented at the Working Group in African Political Economy (WGAPe), University of California, Berkeley, December 5-6.
- Moehler, Devra C. 2008. Tune in to governance: An experimental investigation of radio campaigns in Africa. Paper presented at Field Experiments in Comparative Politics and Policy, University of Manchester, UK, July 1-2.

- Olken, Benjamin A. 2007a. Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy* 115 (2): 200-249.
- Olken, Benjamin A. 2007b. Political institutions and local public goods: Evidence from a field experiment in Indonesia. Poverty Action Lab Working Papers.
- Özler, Berk, Sarah Baird, and Craig McIntosh. n.d. Vulnerable groups component of Tanzania's Social Action Fund (TASAF). Unpublished.
- Paluck, Elizabeth Levy. 2008. Is it better not to talk? A field experiment on talk radio and ethnic relations in Eastern Democratic Republic of Congo. Working paper, Harvard University, Cambridge, MA.
- Paluck, Elizabeth Levy. 2009. Reducing intergroup prejudice and conflict using the media: A field experiment in Rwanda. *Journal of Personality and Social Psychology* 96 (3): 574-87.
- Paluck, Elizabeth Levy. n.d. Information radio in southern Sudan. Unpublished.
- Pandey, Priyanka, Sangeeta Goyal, and Venkatesh Sundararaman. 2008. Community participation in public schools: The impact of information campaigns in three Indian states. Impact Evaluation Series 26, Policy Research Working Paper 4776.
- Savedoff, William D., Ruth Levine, and Nancy Birdsall. 2006. *When will we ever learn? Improving lives through impact evaluation*. Washington, DC: Center for Global Development.
- Vicente, Pedro C. 2007. Is vote buying effective? Evidence from a randomized experiment in West Africa. Economics Series Working Papers 318, University of Oxford, Department of Economics, Oxford, UK.
- Wantchekon, Leonard. 2003. Clientelism and voting behavior: Evidence from a field experiment in Benin. *World Politics* 55 (3): 399-422.
- Wantchekon, Leonard. 2008. Expert information, public deliberation, and electoral support for "good" governance: Experimental evidence from Benin. Paper read at Field Experiments in Comparative Politics and Policy, University of Manchester, UK, July 1-2.