SOCIAL SCIENCE INFLUENCES ON PRACTICE IN LEGISLATIVE DEVELOPMENT: TOWARD BETTER USES OF EVALUATION METHODS

Robert Nakamura and Malcolm Russell-Einhorn*

Paper Prepared for Presentation at the 12th Workshop for Parliamentary Scholars and Parliamentarians, Wroxton, Oxfordshire, 24-25 July, 2015

* Robert Nakamura is Prof. of Political Science Emeritus, Rockefeller College of Public Affairs, University at Albany, SUNY and Malcolm Russell-Einhorn is a Senior Fellow at the Center for Peace Democracy and Development, University of Massachusetts, Boston.

SOCIAL SCIENCE INFLUENCES ON PRACTICE IN LEGISLATIVE DEVELOPMENT: TOWARD BETTER USES OF EVALUATION METHODS

Robert Nakamura and Malcolm Russell-Einhorn

Abstract

This paper is written from the viewpoint of participant-observers in legislative development and focuses on the question of how current "top down" evaluation practices can be reconciled with the newly rediscovered "bottom up" implementation imperatives encountered when policies inevitably have to be adapted to the circumstances in which they are expected to have their effect. The current donor legislative program evaluation model is derived from the confluence of social science ideas about policies as experiments and as scenarios that convert inputs into outcomes, these are combined by aid bureaucracies to provide the means to demonstrate their effectiveness and to exercise managerial control. For the most part, top down policies are judged by their fidelity to initial designs, while bottom up policies are, or can be, evaluated according to the degree of change in behavior they have achieved. The paper discusses the degree to which donor evaluation practices remain largely out of step with the more adaptive policy implementation strategies increasingly advocated and embraced in the development mainstream after several decades of insurgent criticism. It also argues that evaluation regimes that emphasize standardized quantitative or reflexively wish to incorporate costly experimental designs into such projects to answer interesting but often narrow deflect attention and divert resources away from the case study tools and practices that, properly designed and funded, will help not only focus on outcomes, but about how and why the project had influence, which relationships and capacities were developed (mutual adaptation), and what project implementers and partners learned from the engagement. While doing single country case studies well is imperative, there are also opportunities to undertake comparative case studies capable of establishing clearer parameters for effectiveness. While this may seem utterly commonsensical, the current evaluation model is resistant to change because it has incrementally evolved to serve multiple political and bureaucratic purposes and has grown by accretion. And the case studies we favor (usually predominantly qualitative, but often employing mixed methods) are burdened by persistent misconceptions about their value, rigor, and applicability when compared with the unattainable standard of experimental research designs. These misgivings, we argue, miss the point since the "gold standard" of experimental research design is usually inappropriate for deriving useful lessons from legislative development efforts and other institutional reform programs. While this critique is not new, a systematic appreciation of the nature and origins of these problems and tensions is needed, as well as a better understanding of process evaluations and the classic usefulness of the case study method and its potential for more rigor in evaluating progress in legislative development programs.

Introduction

Over the past four decades, social science contributions to policy practice can be viewed as falling into three categories: diagnostic studies of conditions and trends; demonstration research to test possible new programs and policy approaches; and evaluation research to assess the effect and derive lessons from existing programs.[1] In the domestic and international spheres alike, government has borrowed abundantly in all three areas from the policy sciences.[2] A common theme, however, has been the well-known tendency for these innovations to be utilized or adapted for particular operational or bureaucratic purposes in ways that were unintended or grossly oversimplified.[3] Our focus in this paper will be on how social science perspectives have influenced legislative development evaluation models and the consequences of that influence for the implementation of legislative development policies.

This paper is divided into three broad sections. First, we will show how social science work on evaluation has influenced what has morphed into a nearly standard donor agency model for evaluation and program management. The resulting "top down" model—whose hallmark is fidelity to a plan, and the treatment of that plan as analogous to an experimental design--serves multiple and sometimes incompatible purposes when it is used simultaneously to draw summative conclusions about what programs have achieved, make judgments about efficacy of policy designs, and manage implementation of specific efforts. Second, we examine the mismatch between the imperatives of the fidelity model favored by donor agencies and current trends in international institutional reform efforts emphasizing a more flexible "bottom up" approach in which the focus is on the degree of change achieved. Current thinking sees such change as best achieved through mutual adaptation, a process in which plans change in response to found and developing legislative circumstances. Third, we present our ideas for an alternative approach better suited for producing lessons from implemented programs. We favor a mixed method case study that is currently used in the policy sciences but thus far partially neglected by government or denigrated as lacking the authority to draw useful conclusions. The paper will conclude with some suggested research questions that better case study methodologies could illuminate, as well as particular situations where certain method applications (e.g., longer retrospective time frames for evaluation) could and should be encouraged by donors and practitioners alike. Finally, the paper will briefly examine some ideas for comparative case study use, in which broader (but modest) cross-country

---

[1] Richard Nathan, Social Science in Government: The Role of Policy Researchers (Albany: Rockefeller Institute Press, July 2000).

[2] Examples of borrowed innovations include management by objectives/results; the new public management; performance measurement; and strategic planning. Examples of borrowed ideas include rational choice models from economics, components of democratization from political science, and a concern with "best practices" from management and policy sciences. Examples of the utilization of demonstration research findings include the incorporation of the MDRC (Manpower Demonstration Research Corporation) findings in welfare reform (The Personal Responsibility and Work Opportunity Act of 1996).

[3] At last year's Wroxton conference, we attempted to systematically map these bureaucratic distortions or disconnects with the highly complex and politicized world in which legislative development projects operate by identifying a wide range of bureaucratic imperatives that require reform or reconciliation with development imperatives in order to permit effective (and necessarily nimble and adaptive) implementation to take place. These included, among others, (1) the contractual model imperative; (2) the 'outsourcing/labor imperative' in which donors delegate implementation responsibility to others, but then place large numbers of monitoring and reporting requirements on the latter to guard against agency loss; and (3) the evaluation imperative that often privileges onerous process reporting at the expense of more formative or summative evaluation work. *See* Nakamura and Russell-Einhorn, *Improving the Implementation of Legislative Development Programs: Mapping the Imperatives and Circumstances.* Paper delivered at the 11th Workshop of Parliamentary Scholars and Parliamentarians, July 2014.

or intra-regional conclusions about policy implementation and institutional development might be able to be drawn.

Origins of The Assistance Agency Evaluation Model.

At the outset, it is critical to understand how and where the form and language of the social sciences has been incorporated explicitly in bureaucratic requirements for program evaluation and management.

Many major assistance agencies follow an evaluation model that is similar enough to constitute a common paradigm or protocol. Given its ubiquitous quality, it is easy to forget that major elements of that model were once scholarly innovations created for scientific purposes, such as explaining variance or identifying shortcomings in design, rather than for bureaucratic ones, like measuring agency performance. Moreover, many of these currently fused elements were created for separate scholarly purposes and contain their own particular assumptions.

Policies as Testable Hypotheses from Social Psychology

The idea that policies are like hypotheses that can be scientifically tested and refined is actually a relatively new one. Donald Campbell's influential works alerted scholars to the possibilities of a learning society, in which reforms could be considered experiments, and causal inferences could be drawn from variations in practices conceptualized as natural experiments.[4] Such investigations, Campbell and others argued, should follow rigorous protocols paralleling, as closely as possible, those found in experimental psychology, e.g., specifying parameters, establishing controls, etc. These special requirements and qualifications were necessary for discussions of causal relationships, in order to separate them from spurious correlations.[5] *One practical lesson taken from this work was a mistrust of conclusions drawn from observing an unrepresentative sample or focusing on such spurious correlations and failing to institute the requisite controls (the basic concern being that one could prove almost anything if all that was needed was a single example of a correlation).* Case studies were believed to especially suffer from this problem and are often mistrusted by evaluators for that reason. Of course, larger studies are also subject to the same criticisms but are usually better able to deal with them through use of particular designs and controls. Thus, larger scale evaluations were devised as a preferred means for dealing with such problems, and Rossi and others spelled out and scaled up evaluation methods useful for evaluating large scale social programs.[6]

Components of Policy Scenarios

Aaron Wildavsky and others added dimensions to the effort when in the course of their implementation research they conceptualized policies as frameworks in which *policies were seen as scenarios* connecting what policies did (through inputs of resources and activities), to what immediate effects they might have (intended outputs), and how these might affect (outcomes) and ultimately have an impact on society.[7] Wildavsky's preferred method at this time was the case study, because the relationships between these variables were often more readily apparent and more easily operationalized in a single context. These

---

[4] See Donald Campbell, "Reforms as Experiments," American Psychologist, 1969.

[5] Hubert Blalock, Causal Inference in Non-Experimental Research, University of North Carolina Press, 1964.

[6] Peter Rossi, Mark Lipsey, Howard Freeman, Evaluation: A Systematic Approach, 7th Edition, Sage 2004.

[7] For the discussion of the components of policy see Frank Levy, Arnold Meltsner and Aaron Wildavsky, Urban Outcomes, University of California Press, 1974. The Oakland case study presents the policy under study as a set of necessary joint actions, see Jeffrey Pressman and Aaron Wildavsky, Implementation, Third Edition, University of California Press, 1984.

components were later folded into the evaluation model that is more explicitly objective and formal, and that sought to incorporate more quantitative indicators.

From about the 1990s on, these and other ideas about evaluation were adopted by various donors starting with USAID and later extended to others in the same business. USAID itself borrowed from domestic policy agencies such as the National Institute of Education had done a decade earlier. This approach was formalized into an evaluation model that applied to overseas assistance programs, including those in legislative development. Henceforth, programs in a given country were to be evaluated as if they were experiments to achieve specified goals (frequently high-level impacts such as increasing economic growth, reducing corruption, or advancing democracy or good governance) by affecting certain types of legislators' behavior and policy-related actions (outcomes), by increasing their capacity (the outputs of capacity building efforts), and at the beginning of the causal chain, by implementing agency spending on activities (inputs). Results at each level were to be measured by separate indicators, relying insofar as possible on objective and quantitative data, and using comparable measures within each category.

Bureaucratic Frameworks

A look at the results frameworks used by USAID,[8] shows how features parallel those found in DFID logframes,[9] UNDP's Country Program Action Plans, and examples from other agencies. All reflect the influences of scholars like Campbell (treating programs and containing testable hypotheses) and Wildavsky (structuring program elements into related variables) have worked their way into a foundational bureaucratic device for evaluating and monitoring programs.

USAID results frameworks present desired legislative impacts as AOs (Assistance Objectives) such as "Legislatures oversee the executive branch and fulfill their representative and law making functions in and independent manner." These objectives are to be achieved through IRs (Intermediate Results), the equivalent of Wildavsky's outcomes. Intermediate Results in the legislative development field have included: "the law provides for the legislature's independence…," "internal operations….are transparent and operate effectively," "legislators and staff demonstrate increased effectiveness…,." "increased and more effective interaction between legislators and citizens...." These results were to be operationalized into indicators mandated to be separate for each level, to safeguard against double counting and maintain the idea that one was supposed to produce changes in the other. The idea was to keep distinct the relationship between the variables (paralleling the scholarly notion of independent and dependent variables) so that there would be a basis for asserting a causal relationship and through that, apportion appropriate credit to the agency.

Similarities can be found in DFID logframes. The DFID Logframe for the Nigeria Business Case, for example, was organized by impact ("Strengthened Democratic Governance in Nigeria"), measured by an impact indicators (such as the Ibrahim Index Score for Participation), to be served by outcomes (such as

[8] This description of the Legislative Results Framework is drawn from David Olson and Lynn Carter, *Indicator Gap Analysis: Legislative Strengthening*, Report to USAID from Management Services International, September 2010.

[9] See Greg Power, "The Logframe and the Beautiful Game: Project Logic vs. Football Logic," Politically Agile Programming Paper 1, Global Partners Governance.

acceptance of election results, etc.), which are related to inputs (which include what Wildavsky called outputs) in the form of program activities.[10]

The UNDP Country Program Action Plan for Bangladesh, meanwhile, also bears characteristics closely resembling the USAID and DFID frameworks: outcomes in the form of increasing legislative capacity and oversight, and reforming institutional practices, were to be affected by outputs (better skills or plans) produced through UNDP program supported inputs (training, etc.).[11]

In the end, what was a test of policy ideas through examining natural experiments in Campbell, and examining the links between elements of policy in case studies has become a device for establishing the credit due to a bureaucratic agency in carrying out a governmental assistance program.

The Evaluation Model as Management Tool: The Principal Agent Problem and Fidelity to Plans

Another legacy from scholarship has worked its way into the framework from rational choice modeling: concern with the principal-agent problem.[12] The principal agent problem is encountered whenever a principal delegates to an agent responsibilities to perform tasks that principals either cannot or choose not to perform for themselves. The goals of principals may be separate from those of agents, and insofar as agents can act independently of principals, the possibilities for "agency loss," the diversion of the principal's resources to benefit agents, increases.

While a problem in both the public and private sectors, and within the public sector between political masters and bureaucrats, agency loss is a particular concern of legislative assistance donors for two reasons. First, the donors operate primarily or almost exclusively as funding agencies. Which means they achieve their goals through the spending of "money with strings." Money diverted by agents or target populations to other uses and other goals thus directly undermines the usefulness of the donor's main resource. As a corollary, these agencies rarely run legislative assistance projects themselves, depending instead on "contracting-out" for these services.[13] Typically, principals lack the time and/or expertise to manage agents and depend instead on devices to perform these tasks for them.

The device donors often choose to deal with the principal-agent problem is a combination of the assistance framework and the workplan (or its equivalent), which spell out in detail both the activities for which money can be spent and performance measures for monitoring the quality of contractor efforts. In Wildavsky's terms, workplans limit the use of resources for pre-ordained inputs and assess the effectiveness of implementers through specified (and usually quantified) outputs and outcomes.

Typically, these frameworks and plans are laid out in detail at the outset of project implementation, when the least is known. And the contracted agents are primarily assessed by their degree of fidelity to fulfilling the tasks laid out and in delivering on the results indicators specified. Fidelity in psychology

---

[10] For a description of these components, *see* DFID, "How to Note," Practice Paper giving guidance on how to write a business case.

[11] See Robert Nakamura and Carl DeFaria, Final Evaluation: Improving Democracy through Parliamentary Development, written for UNDP Bangladesh, June 15, 2014.

[12] The most influential statement of this problem comes from Terry Moe's 1985 article, for a more recent discussion see his "Political Control and the Power of the Agent," Journal of Economics*, Law and Organization*, 2006.

[13] This can take a variety of forms: USAID and DFID often use contracts with organizations, they and others may also fund civil society groups, and UNDP relies on chief technical advisors who are contracted for fixed periods to run their operations.

experiments basically translates into strictly following a prescribed protocol so that the hypothesis being tested is accurately translated into action. In bureaucracies, little testing of the central impact propositions is possible or desired. Rather, impacts and their relationships to outcomes is fundamentally assumed or finessed, and the performance focus is more on using the relationship of inputs to outputs as a management tool. *Fidelity is a device intended to reduce agency loss, and usually unsuited to the causal weight ascribed to it by overly schematic and linear logical frameworks in the policy and institutional reform arena.*

Problems with the Derived Evaluation-Implementation Model

How did these disparate strains and traditions in evaluation research, implementation studies, and political economy come together in the form of a framework for assessing and managing bureaucratic performance? One possible explanation is provided by John Kingdon's confluence model, in which the stream of ideas, the political needs of actors, the need to do something about a problem come together at the same time.[14] The connections are agenda-driven by the necessity to act, so ideas are brought together for political and bureaucratic reasons.

Such a confluence appears to have occurred in government in the early 1990s when social scientists' ideas about program evaluation were being widely noticed and acted upon: witness the spread of program budgeting and the linking of bureaucracies with measureable performance goals, the passage of legal requirements for sunset laws requiring examination of the worth of activities, and various laws mandating evaluation for federal programs. Principal-agent concerns were also coming to a head in defense and other areas where over-spending scandals resulted in a move toward more rigid contracting and procurement procedures. At this same time, aid agencies were converting themselves from direct assistance agencies (food and commodity drops) into the managers of large-scale institutional reform programs and efforts to improve the capacity of governments. But while the scope of their efforts was increasing with rising demand in those areas, the anti-government movement was constraining the growth of these agencies, leading them to *contract out for necessary tasks*. And the same mistrust of government was pushing the legislative masters of these agencies to demand greater accountability and evidence of efficacy as a condition of funding.

The result was the agency evaluation model that combines the idea of policies as experiments, with scenarios that connect agency inputs to societal impacts, to produce a tool for claiming credit and for reducing agency loss.

The Kingdon explanation suggests that there will be difficulties when solutions are chosen for political reasons rather than their effectiveness in addressing actual social and service delivery problems. We discuss a few of the problems with the agency evaluation and management model.

1. ***Diagnosis and Attribution.*** Sometimes these goals are at odds. Using evaluation to learn from implementing policies and using evaluation as device for assigning credit are often incompatible when done simultaneously.[15] How, for example, is finding a problem treated? *Diagnosis uses mistakes as a learning tool and therefore benefits from finding shortcomings.* Wildavsky said

[14] See John Kingdon, Agendas, Alternatives and Public Policies, Update Edition, Longman 2010. , see also Michael Cohen, James March and Johan Olson, "A Garbage Can Model of Organizational Choice," Administrative Science Quarterly, March, 1972.
[15] This is a criticism that has dogged "No Child Left Behind" and the use of education test scores.

that the problem was not making mistakes, but making them fast enough so that we can learn from them to improve practices. [16] *In the context of assigning credit or fidelity to plans, finding shortcomings may be a step toward invoking sanctions* against either the contractor or the aid agency. Not surprisingly, there is a bias among both funders and implementers (principals and agents) toward finding success at the expense of identifying lessons. Witness the contrast between the pessimistic conclusion of outside observers about the efficacy of assistance efforts[17] and the frequent findings of success in program evaluations. An example of this problem is found in the frustration of recent efforts to mine the USAID data base of project evaluations for a fuller and more illuminating set of lessons.

2. ***Stretch and Overreach.*** The connecting of modest inputs to grand impacts (or even to shorter range outcomes) through the implicit scenarios embraced at project inception (or earlier) lead planners to go far beyond what is known. They are, after all, under the gun to show big results from modest resources and such scenarios are useful for describing a path for doing precisely that. This level of ambition obviously did not characterize the evaluation work from which the model is derived. It is useful to remember that the initial work in evaluation from Campbell dealt with relatively concrete policies and plausible relationships (e.g. stricter traffic enforcement and highway accidents), and even where the goals were ambitious such as in Wildavsky's Oakland study of efforts to use economic development to reduce hardcore unemployment, the focus of his case study based evaluation was primarily on the reasonable goal of putting plans into practice rather than the more remote concern with achieving of bigger results. But the aid frameworks discussed often connect concrete efforts (such as training sessions, study tours, etc.) through what are plausible outputs to more ever more remote and often highly contingent outcomes. And the hypothesized link between these outcomes and impacts are often not well grounded in scholarly consensus. Examples of this last point include the achievability and desirability of goals like disciplined parties and democracy,[18] or in the links between more effective democracy, economic growth and equitable distribution of benefits.[19]
3. ***Reliance on problematic impact indicators.*** The very metrics that donors use to measure outcomes such as functioning legislature are themselves in question. While non-functioning legislatures are easy enough to identify by what they lack or don't do, there is simply no consensus on what constitutes a functional legislature.[20] The available definitions and indicators have been devised for a variety of purposes and are often fall short used for donor's multiple purposes.[21] Furthermore the most frequently used impact variables, such as the quality of democracy or governance, are themselves problematic. Current definitions of democracy, for example, suffer from a variety of shortcomings in definition, precision, sources

[16] Aaron Wildavsky, Speaking Truth to Power: The Art and Craft of Policy Analysis, Little Brown, 1979.

[17] See Lant Pritchett, Michael Woolcock, Matt Andrews, Capability Traps?: The Mechanisms of Persistent Implementation Failure, May 2010.

[18] Much of the party support effort of donors is based in a vision of a "responsible two party model" which has long been disputed as a description of effective democracy and as an attainable goal. In 1950, the American Political Science Association advocated a more disciplined, ideological party system, that view was disputed in Evron Kirkpatrick, "Toward a Responsible Two Party System: Political Science, Policy Science or Pseudo Science," American Political Science Review (December, 1971). Recent changes in the American party system has stimulated further controversy about whether such systems are desireable, William Galston, "Can a Polarized American Party System Be Healthy?," Brookings Issues in Governance Studies Paper, April 2010.

[19] See, for example, Michael Ross, Is "Democracy Good for the Poor?," American Journal of Political Science, Oct. 2006.

[20] See Robert Nakamura, "Legislative Indicators: Measuring How We Are Doing Depends On What We Want To Do," forthcoming.

[21] See Olson and Carter, op.cit.

and coverage, coding, aggregation, and the failure to institute reliability and validity test.[22] They are "especially inadequate for smaller changes in democracy" that are targeted by development programs.[23]

4. ***Time and time horizons.*** In evaluations intended to draw lessons about policies, time frames are chosen for technical reasons such as allowing adequate time to effect changes and produce results. By contrast, evaluations designed to apportion credit to a program or agency use time frames that generally simply coincide with the involvement of the actors commissioning the efforts. Often program-determined time frames in legislative development are short—two or three year commitments—and not synchronized with the requirements of the target populations. For example, many legislative programs start when donors have the money and can conclude agreements on schedules that ignore election cycles that produce new parliaments. And the adequacy of the short program determined time frames seem especially questionable when compared with the longer time frames used by scholars when they study how legislatures have institutionalized over time.[24] Similar findings about the importance of longer time frames for evaluation are found in policy implementation research which has emphasized the reinforcing or cumulative effects of many policies intended to achieve the same goal over time.[25]

2. The Model and Current Development Thinking

We noted that fidelity in the use of program resources was a safeguard to protect the donor's major resource from diversion. In addition, it has served as a management tool for keeping contracted agents from wandering too far from principal's goals during implementation. But the requirement of fidelity to plans and achieving pre-determined goals is at odds with what many implementation scholars and practitioners have found to be effective. Implementation research long ago examined the problems of implementing policies in uncertain environments. Many found that greater flexibility – the need to adapt policies to given environments through mutual adaptation—was a necessary component of effective program management.[26]

A similar critical perspective extending back at least two decades can also be found in the development literature. These older viewpoints along with those of implementation researchers are now being re-discovered in the context of a new wave of development thinking. Over the past seven or so years, there has been a virtual cascade of articles and forums decrying the poor state of much of the international development enterprise, taking particular aim at 'projectized,' externally-led (at least in design terms) development that is wedded to highly prescriptive designs, specified outcomes and

---

[22] Michael Coppredge, John Gerring et al., Conceptualizing and Measuring Democracy: A New Approach," Research Notes, June 2011.

[23] Coppedge, Gerring et al, A New Approach to Conceptualizing and Measuring Democracy, Paper presented at the 3rd International Conference on Democracy as Idea and Practice, Oslo, January 2012. Page 9.

[24] See for example, Nelson Polsby, "The Institutionalization of the US House of Representatives," American Political Science Review, August 1968 and Samuel P. Huntington, "The Congressional Response to the Twentieth Century" in David Truman, ed., Congress and America's Future, American Assembly, 1973.

[25] See, for example, Martin Levin and Barbara Fermin, The Political Hand, Pergamon, 1985. See chapter on advocacy coalitions in Paul Sabatier and Christopher Weibel, eds., Theories of the Policy Process, Westview, 2014.

[26] Milbrey McLaughlin reported on the Rand Change Agent Study in , "Change as Mutual Adaptation: Change in Classroom Organization, in David Flinders and Stephen Thornton eds. Curriculum Studies Reader, Routledge 2008. Michael Lipsky, "Toward a Theory of Street Level Bureaucracy," discussion paper, University of Wisconsin, 1969. Robert Behn, "Leadership Counts," Harvard University Press, 1998. Eugene Bardach, The Implementation Game, MIT Studies in Public Policy, 1977. Robert Nakamura and Frank Smallwood, The Politics of Policy Implementation, St. Martins, 1980.

outputs, and a technocratic mindset that eschews real engagement in local politics. The recent literature, largely if loosely captured within the "Doing Development Differently" (DDD) movement, advocates for a much more flexible, adaptable, politically sensitive approach to program implementation that is abuzz with references to 'managing complexity,'[27] 'working politically,'[28] and assuming a 'problem-driven' approach that works over longer periods of time to seek 'iterative' solutions.[29]

There are many important and interesting insights and lessons learned being synthesized in this new community of constructive critics, who have brought fresh attention to the importance of bringing a systems thinking orientation to policy reform work, and to the issues of how distributed capacities, networks, and collective action problems – not to mention leadership – affect policy implementation. One has the impression, however, that many of these insights amount to lessons learned and unlearned—or surely in many cases never learned—despite many antecedent efforts to explain and improve obvious shortcomings in policy implementation. Over nearly a decade of work, USAID's *Implementing Policy Change* research initiative identified a number of critical ingredients for successful reform program efforts, including the importance of understanding reform as a contested arena with winners and losers, employing political mapping tools to better understand the potential stances of various stakeholders on particular issues, the vital tasks of mobilization and coalition-building, and the use of a wide range of deliberative, participatory forums to engage in design and implementation problem-solving.[30] Meanwhile, as far back as the 1980s and early 1990s, shrewd and practical observers such as Dennis Rondinelli ruefully perceived the emergence of overly rigid and technocratic bureaucratic routines in development work and advocated eloquently for greater use of adaptive strategies, flexible management structures, better networked communication in contexts that were clearly complex and posed huge problems of uncertainty to reformers.[31]

Insofar as some of this knowledge has in fact been absorbed by astute and conscientious development professionals over the past 25 years—especially those who work in governance generally and legislative development in particular, some of the new discussions and critical perspectives have a straw man quality about them. It is widely understood that successful legislative development projects do work in an iterative, adaptive, and politically astute fashion, and that local leadership, relationship-building, and networked implementation have been central to such success. The same is true of encouraging greater deliberation among key reform stakeholders during both program design and implementation, and that local solutions are at the heart of sound reform approaches.

---

[27] Jones, H., 2011. *Taking Responsibility for Complexity: How Implementation Can Achieve Results in the Face of Complex Problems,* Working Paper 330 (London: Overseas Development Institute). More recently, *see* Root, Hilton, H. Jones, L. and L. Wild, 2015, *Managing Complexity and Uncertainty in Development Policy and Practice*. London: ODI.

[28] Booth, D and Sue Unsworth, 2014. *Politically Smart, Locally-Led Development.* Discussion Paper (London: Overseas Development Institute).

[29] *See, e.g.*, Andrews, M. and L. Pritchett, 2012. "Escaping Capability Traps Through Problem-Driven Iterative Adaptation," *Center for Global Development Working Paper* 299.

[30] IPC's work generated dozens of publications and guides, many of whose insights were synthesized in Brinkerhoff, D. and Ben Crosby, 2002, *Managing Policy Change: Concepts and Tools for Decision-Makers in Developing and Transition Countries.*

[31] Rondinelli, D., 1993. Development Projects as Policy Experiments: An Adaptive Approach to Development Administration (2nd ed.)(Boulder: Lynne Rienner).

Interestingly, among the arguable straw men cropping up in the DDD discourse is the notion that project management and accountability structures are not flexible—when in practice, at least in legislative strengthening and other governance projects, they usually are. In fact, frequently, *if not routinely,* many of these facially onerous requirements are bent to accommodate such factors as changed political circumstances, the use of multiple entry points to explore and exploit reform opportunities, and iterative learning and problem-solving generally. In fact, there is constant negotiation and bargaining between and among donor and implementer staff regarding fidelity to workplans (what we have previously referred to as the 'reconciliation' imperative')[32] as both sets of actors attempt to balance, as adroitly and commonsensically as they can, certain predetermined program design features and reporting requirements with evolving (if unanticipated) pathways and indicia of success. While overly simplistic and linear workplan designs often extract a real toll on projects in terms of transaction costs (and program staff stress and distraction), it is remarkable how the new DDD discussions tend to see many, if not most projects as irrevocably straitjacketed when in fact this inevitable reconciliation with reality on governance projects (accomplished formally or informally) is commonplace.[33]

Regardless of the degree to which – and the reasons why – some of the new lessons learned are less 'new' than is generally appreciated,[34] the development community (and especially those who work in governance) are in fact asking hard and important questions about certain programming fundamentals, particularly in light of 'sticky' bureaucratic path dependencies. For example, it's fine to advocate for a more problem-oriented and iterative approach to programming, but even five year projects may be too short and funding streams too fickle to empower policy entrepreneurs and significant program experimentation ('failing fast and cheap' to generate new learning, as many DDD partisans advocate, may simply be practically and politically infeasible). And the question of who ultimately creates and 'owns' the definition of the 'problem' may still be murky, even when a program is ably and locally led. Finally, acknowledging the true extent of contextual complexity creates new dilemmas of when is enough knowledge of an ever-changing political, economic, and social landscape enough? Many thoughtful practitioners have raised the possibility, in the future, of multi-agent social simulations to identify opportunities and risks of particular policy options. If government donors are unable to grapple with some of these more interesting questions in the years ahead, it may be left to more entrepreneurial private funders to embrace these challenges.

Program Evaluation: Out of Step?

In spite of the continuing formidable obstacles to mainstreaming many of the DDD precepts in donor agencies in the near term, substantial changes in program design, procurement, and program management are well underway in many bilateral donor agencies, most notably in two of the largest—

---

[32] *See* Nakamura and Russell-Einhorn, *Improving the Implementation of Legislative Development Programs: Mapping the Imperatives and Circumstances.*

[33] This is of course more the case with grant or cooperative agreement vehicles, but it is a barely-disguised secret that many of the more onerous contractual workplan modification and results reporting requirements are routinely finessed in various ways.

[34] There may be many reasons for this, including significant declines in funding for governance programming (affecting cadres in major donor organizations and an associated hollowing-out of expertise within donor and implementer organizations alike), inadequate training and mentoring of new staff in many development organizations regarding lessons learned, and of course growing bureaucratic reporting and coordination demands on staff that leave little time for reflection.

USAID and DFID.[35] In one crucial area, however—program monitoring and evaluation—agencies have seemed to be oddly out of step. Even as the need for program flexibility has grown, donors and implementers have not only been subject to *increasing* standardized performance measurement demands generally (including continuing requirements for standardized indicators that follow the mechanical logframe methods discussed in the previous section of this paper), but to *new and more elaborate expectations about evaluation methods.* Such expectations go beyond the need for better, more informed monitoring and learning within projects, or the need for better independent mid-line, end-line, or retrospective evaluation—all of which are genuinely needed in many areas of development assistance, especially governance-related programs. And they go beyond some of the more onerous, one-size-fits-all performance measurement activities packed with multiple, often conflicting values and expectations (among them that outcomes can be easily and clearly measured) criticized a decade ago by Beryl Radin.[36]

Rather, for the most part, these relatively new expectations revolve around a desire to demonstrate program success through rigorous experimental evaluation methodologies that can not only generally seek to assign agency credit or contribution to an outcome, but rigorously isolate and attribute outcomes and impact to particular interventions and associated independent variables. The simplistic infatuation with 'gold standard' RCT evaluations (still holding sway among many top officials in the US government (at least in many parts of the State Department, USAID, and the Millennium Challenge Corporation) thoroughly misses that method's narrow applicability to most development interventions and its hefty cost). While below the radar, most implementers and line development officers have in fact drastically scaled back on the numbers and scope of more ambitious evaluation efforts, ideological commitment to the gold standard (and to multi-variate regression analysis—still very popular at the World Bank)[37] remains strong not only among many with backgrounds in the sciences (esp. those working in health and agriculture) but many politicians and senior donor officials impatient with the lack of rigor in existing evaluations (a continuing problem), particularly those purporting to convey the results of institutional strengthening and governance-related programs.[38] Legislative development programs, though modest in number, are obviously among these.

The desire to incorporate randomized controlled trial (RCT)-type designs into all kinds of development assistance programs—including some decentralization programs, ostensibly due to the potential

---

[35] USAID launched significant procurement reforms several years ago and has been moving to shift a larger proportion of programming to grants from contracts. DFID has adopted a new program of "Smart Rules" to simplify a broad range of management functions to encourage greater programming flexibility.

[36] Radin, B., 2006. *Challenging the Performance Movement: Accountability, Complexity, and Democratic Values* (Washington, DC: Georgetown University Press).

[37] Certainly this kind of statistical analysis can provide evidence of the strength of causal relationships across many cases in a sample (while still examining the role of many potential causal variables), but as many students of development have noted, such methods provide a very limited understanding, if any, of the actual causal mechanisms that produced the outcome in individual cases. They also frequently require the use of crude proxies for variables that may similarly obscure the processes actually at work.

[38] Another frustrated constituency were some of the private foundations (most notably the Gates Foundation) whose leadership strongly believed that more demanding research designs could yield better program impact and generalizable knowledge. At the same time, it must be noted that the Gates Foundation also commissioned ODI researchers to use rigorous case study methods to study sector specific service delivery outcomes, focusing on the role that leadership, institutions, policies and foreign actors played in influencing development outcomes. Overseas Development Institute, 2011. *Mapping Progress, Evidence for New Development Outlook*. London: ODI.

availability of subjects with a Large-N and a presumed (but invariably mistaken) ability to reasonably control for many key variables – continues to epitomize an even newer embrace of academic research rigor and desire to demonstrate results to politicians and a pubic perennially disillusioned by most foreign assistance work.

Recall that we alluded above to the origins of this view with thinkers like Donald Campbell. Such investigations, Campbell and others argued, should follow rigorous protocols paralleling, as closely as possible those found in experimental psychology: e.g., specifying parameters, establishing controls, etc. Moreover, special requirements and qualifications were necessary for discussions of relationships, in order to separate them from possibly spurious correlations. Such controls were more practical when using larger data sets and case studies were correspondingly mistrusted for the same reason.

The problem, again, is that a donor (or practitioner) preference based a priori on concerns about methodological rigor privileges an approach that may well not be suited for the task at hand and therefore breaks a scholarly rule to choose methods appropriate for what you wish to find out. These larger-scale evaluations with higher sample sizes are usually ill-suited to the kind of highly complex and uncertain environments that characterize most development assistance work and that have inspired much of the DDD movement's thinking. As many have noted for years, most recently Rachel Kleinfeld,[39] they are particularly ill-suited to social and political reforms in the governance arena that are invariably contested, have multiple or ill-defined goals, and are marked by many stops and starts and iterative adaptation.[40] RCTs may be highly probative of the effect of a given intervention on a particular outcome under highly controlled or stable parameters (by definition, often highly technocratic and artificially apolitical ones), and may even point the way to optimal solutions within those parameters, but they are usually unhelpful in explaining the often complex processes that produce an outcome in highly uncertain environments.

The fact is that not only are these evaluations expensive and of dubious value in showing how things can be done elsewhere under different conditions, they crowd out funding and attention paid to formative evaluations and learning (which all the aid agencies profess to be encouraging). They generally can show that something worked, but are hard-pressed to show precisely that the processes worked in a certain way (i.e., they can't easily demonstrate that a process can be replicated elsewhere).

3. Toward More Useful Evaluations

In reality, evaluation methods need to move in tandem with the assistance community's new focus on flexibility and experimentation—as opposed to improvements in meeting agreed-upon standards or goals—where both developmental paths and the end objectives are invariably evolving. This suggests the use of mixed methods[41] to explore a wide range of relevant, complex factors affecting outcomes,

---

[39] Kleinfeld, R., 2015. *Improving Development Aid Design and Evaluation: Plan for Sailboats, Not Trains.* Washington: Carnegie Endowment for International Peace.

[40] See Roche C. and L. Kelly, *Monitoring and Evaluation When Politics Matter,* Development Leadership Program, Background Paper no. 12.

[41] A mixed methods approach is also advocated by those seeking better measures of democracy. See Coppedge, Gerring et al, op.cit. above

and to help illuminate contradictions and hidden biases. These mixed methods (including ample demographic and survey information) can also be employed to properly identify and query key stakeholders and avoid or minimize blind spots or power differentials in the definition of problems and solutions.[42] Mixed methods can thus be used to help strengthen both summative evaluations as well as process, or process evaluations. With so much pressure for 'results,' formative evaluations have not been given nearly enough attention by donor organizations, despite their new emphasis on learning and adaptation. Summative evaluations, as we have seen, have been in thrall to larger scale models that seek highly generalizable findings that may be both elusive and cost-ineffective. Here, we briefly describe, first, some of the benefits that may be derived in the legislative development arena from innovative process evaluation work, and then proceed to discuss the importance of reviving the use of case study methodologies to help evaluate some of the more significant outcomes from various kinds of legislative development programs.

Using Innovative Process Evaluation Methods

Using mixed methods, a number of innovative process evaluation approaches have emerged in the past decade and a half that are more attuned to today's adaptive learning and iterative implementation modalities, although they have only recently been seriously explored by donor agencies and implementers – and are still far from reaching the mainstream of development assistance work. Going by different names and with particular methodologies, but serving overlapping purposes, so-called developmental evaluation, transformative evaluation, and outcome mapping are increasingly being used as collaborative, participatory evaluation vehicles. They are designed not only to create a better body of knowledge and help evaluation be more realistic, useful, and contextual, but to encourage 'process' changes in the development enterprise itself to be carefully understood and tracked—particularly changes in the behaviors of those with whom a project interacts directly and has influence as partners.

Consequently, changes in partner behavior—both implementing partners and beneficiary partners before, during and after a project—*are among the most important results to be observed in the evaluation effort.* These behaviors can include issues of collective learning, increased ownership of interventions, and better engagement of mutual support for further behavioral change. And since success on many types of development projects (especially those in the legislative strengthening field) depends precisely on behavior or social change in key groups or networks, data on changes in relevant behaviors are often equally or more important than more classically 'relevant' impact indicators. Instead of being preoccupied with attribution, the emphasis is on tangible changes in the actions and behaviors of the actors involved.[43] In the legislative development arena, these can include measurable changes in organizational arrangements, relationships and networks within a legislature and extending outward to individuals and groups in other parts of government and civil society, as well as improvements in relevant skills and work habits among members of these communities. It can also embrace, to a point,

---

[42] Mertens, D., 2007. "Transformative Paradigm: Mixed Methods and Social Justice," *Journal of Mixed Methods Research 2007*, 1:212.

[43] Smutylo, T., 2005. Outcome Mapping: A Method for Tracking Behavioural Changes in Development Programs. ILAC Brief 7. Yet another pragmatic, process-oriented evaluation methodology is Utilization-Focused Evaluation (UFE), developed by Michael Quinn Patton, which focuses principally on an evaluation's usefulness to its intended users and seeks to ensure that these intended uses of the evaluation by the primary intended users guide all other decisions that are made about the evaluation process. Thus, rather than focus on general and abstract users and uses, UFE is focused on real and specific users and uses. *See also* Patton, M., 2007. "Process Use as Usefulism," *New Directions in Evaluation* 116: 99-112.

certain tangible outputs (e.g., the quality of research, reports, or legislative drafts). The idea is that the main focus of attention should be on the places where the project has the most *influence* – which often is on the thinking and behaviors of the local partners – and not necessarily on certain outcomes or impact, which remain (or at least should remain) the responsibility of those partners.[44]In addition to these uses, many of these process uses can be profitably applied to the work of development organizations carrying out the legislative assistance work. Such evaluation methods can, and do, help such organizations with their own organizational culture and performance, opening up new perspectives on the nature of the project and specific interventions, deepening commitment and shared learning/understandings, and focusing attention on evaluation priorities and methodological adjustments, to name just a few possible benefits.

All of these innovative process evaluation methods are highly supportive of the more exploratory kinds of program design and implementation being explored by development innovators. They can comfortably coexist with a variety of traditional evaluation methodologies imposed on implementers as a matter of a funder's results framework accountability requirements. Due to their focus on evaluator and user ownership, they also help partner organizations appreciate and be more conscious of their evaluation work not as a separate enterprise from project implementation as such, but as central to their programs, including the tasks of fostering greater sustainability and local capacity to work iteratively and with greater numbers of policy options and choices.[45]

A Needed Re-Appreciation of the Value of Rigorous Case Studies

Often methods discussions start out with a low common denominator view of case studies and an idealized conception of the alternatives. If we use academic literature in public policy as a measure of acceptable rigor, articles based on case studies continue to be an important component. And comparative case studies are an increasingly important way of evaluating government programs intended to spur change in diverse environments such as revenue sharing, educational innovation, and medical care.[46]

Case study approaches have been overshadowed because of the exaggerated expectations and confidence placed by some donors in such tools as RCTs and multi-variate regression analysis regressions to demonstrate 'impact' (although these donors often do not understand their uses). Despite this, case studies still have a central role to play in investigating not just 'what works' in development programs contextually, but also why how and why. They are also very good at teasing out the unintended consequences and blind alleys encountered by well-intentioned and even very well-designed programs (which still have a tendency, as do many evaluations, to steer toward pre-conceived causal pathways and anticipated results). And they excel as a means for situating what happened in the context of time and circumstances.

---

[44] Carden, F., 2009. Knowledge to Policy: Making the Most of Development Research (London: Sage Publications), p. 177.
[45] Carden, Knowledge to Policy, p. 192.
[46] Richard Nathan and Alice Rivlin direct the Affordable Care Act Implementation Research Network which employs a field network approach studying ACA implementation across the states. A similar approach was taken to study welfare reform, see Thomas Gais, Richard Nathan and Irene Lurie, Implementation of the Personal Responsibility Act of 1996. Nathan also conducted field network research on federal revenue sharing. Many of the resulting publications are available through the Nelson Rockefeller Institute, SUNY, Albany, New York.

An important feature of case studies is that they are focused on understanding events in the context, and often address not just with what has happened, but why. Especially when employing mixed methods, case studies remain a serious research method for rigorous and fair presentation of data on the kinds of questions of interest to many in the development community now; namely, how do such factors as leadership, crisis, collaboration, advocacy, institutionalization and formalization, relationship-building, local ownership assumption, and network formation and information flows affect social and political change, and why. And particularly the role of culture in all of the foregoing.[47]

Insofar as case studies themselves can be somewhat imprecisely defined, lacking uniformity as a group and embracing multiple possible research designs, their virtues can be obscured by the difficulties they pose for aggregating findings in the form of general conclusions. This diversity is also a source of their strength, as they use many tools– structured and semi-structured interviews, deep observation, participant observation, focus groups, surveys, social media sampling, and many different kinds of quantitative data – not available in alternative approaches. As Martin Levin observed, implementation case studies usually have the virtue of getting the story straight—of telling the story of what has happened in its fullness and complexity.

Case studies fundamentally seek to understand the basic contributors to progress in development, asking questions and supplying analysis about how the sequencing of critical decisions and events and problems and constraints affect development outcomes. In this way they are precisely oriented to the multi-causal, non-sequential nature of political, economic, and social change, and they do so through a concentrated "analysis of a single or a small number of units, where the researcher's goal is to understand a larger class of similar units."[48] Judicious but rapid use of significant amounts of macro- and micro-level data and "thick description" can also help researchers and implementers refine key assertions and hypotheses, modify data collection tools and add additional data sources to provide an even more in-depth understanding of key issues at stake in any given development intervention, further serving the needs of problem-driven, iterative adaptation. All of these advantages tend to distinguish case studies generally from other methods, which seek to understand what happens with many, if not all, similarly situated interventions, sites, or populations.

Reinvigorating the use of Case Studies to Improve Learning in the Legislative Development Field

Despite the recent pressures placed on legislative development practitioners to utilize potentially unsuitable quasi-experimental evaluation methods to analyze reform initiatives with a strong institutional development thrust, there is a long and well-respected tradition of quality case study work that provides a 'usable past' for further meaningful, contextual learning in the field. Case study methodologies are especially relevant and useful given the highly politicized nature of legislatures, election turnover, complex relationships between and among leadership officials and regular members,

---

[47] Flyberg, 2001. Making Social Science Matter: Why Social Inquiry Fails and How it Can Succeed Again. Cambridge: Cambridge University Press.

[48] Seawright J. and J. Gerring, 2008. "Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options," *Political Research Quarterly* 61: 294-308.

among members and professional staff, and between legislative personnel and representatives of other public bodies and civil society.

We offer a few examples of case studies that have focused on legislative development objectives and the conditions that have promoted their achievement. Each treats results as end states shaped by what happened rather than a pre-determined standard to be achieved. More importantly, each presents the legislative environment as a politicized place, in which actors with multiple preferences interact and change in response to circumstances. In this world, the often more static assumptions made in larger scale studies and in donor frameworks are less appropriate for understanding what is happening on the ground in particular places.

Buy-In and Ownership

An important goal of legislative development projects is getting legislators to better understand and adopt institutional development goals as their own and therefore promote the sustainability of changes and the likelihood that they will be utilized. Too often terms like ownership and buy-in are thought of as the products of formal negotiated agreements with governments, pledges of loyalty to abstract goals, etc. They are, as case studies have revealed more often the results of various processes, adjustments, and shifts of thinking, and usually come from energizing small groups of reformers rather than whole institutions. Joel Barkan's case studies of African legislatures and our own recent work on Kenya are examples of how a more fine-grained examination of circumstances, places, actors, and events can yield useful lessons about what is required and how to proceed.[49] Both, it should be noted, employed mixed methods, including judicious use of significant quantitative changes in various indicators of institutional capacity and maturation.

What Works?

While legislative development projects and other governmental organizational or policy reform efforts are about shaping grand institutional systems of relationships, the more modest and plausible goal should be creating smaller scale functioning relationships that can, with luck, be scaled up over time. Greg Power has likened legislative assistance efforts to creating a small business in the expectation of reforming the economy. Matt Andrews has pointed out that statutory changes to create functioning relationships often fall short because the behaviors of those in institutions do not change in the expected direction. Case studies of functioning relationships on a small scale are a means of learning about how such connections can be made to work and produce desirable results as well as providing subsequent policy planners with lessons they can use in future plans. Again, these are tales of influence, not impact. Andrews own studies of the experiences of particular efforts to institute some best practice approaches provide knowledge about important sources of resistance.[50] Other work on the success of small scale oversight programs in which legislative committees, audit agencies, and civil society have functioned to produce modest changes provide insight into why such relationships sometimes work.[51]

---

[49] See Joel Barkan and Fred Matiangi, "Kenya's Tortuous Path to Legislative Development" in Barkan et al African Legislatures forthcoming. Robert Nakamura, Heather Senecal, and Andrea Wolfe, "Creating and Implementing a Kenya Model of Parliamentary Development," Paper Presented at the 12th Workshop for Parliamentary Scholars and Parliamentarians, Wroxton, July 2015.

[50] Matt Andrews, Limits of Institutional Reform, Cambridge, 2013..

[51] Robert Nakamura, "Effective Oversight and Iron Triangles: Toward a Better Understanding of Challenges or Lessons from BiH and Turkey, paper presented at the 2014 International Political Science Association Meeting, Montreal. See also Robert

Management Strategies

An important continuing question in legislative development is how much discretion should donor agencies allocate to their agents. And how should those allocations be managed and the results judged? These questions are best discussed in context rather than in the abstract and in the aggregate. The experience-based work of Greg Power and others provides both a framework for understanding the need for discretion in working in the Iraqi parliament, but also offers some criteria for judging such efforts (e.g., the relative resilience of changes).[52] Recent work on Kenya focuses on the need for discretion in managing legislative development over a decade of changing parliaments.[53] Since implementers need discretion to work in legislative environments, and donor agencies have to be concerned with how their resources are managed to produce goals that they value, the issue of how to manage is a crucial one. These and other insights, gleaned from case studies with various degrees of rigor, provide potential lessons for how to draw and justify the line between an agency granting necessary discretion and giving up control.

Comparative Case and Field Network Studies

Case studies are useful and they can and should be taken more seriously by donors—who should model good practices in case study methodology by incorporating adequate funding for such studies into independent evaluation and learning budgets and circulating more widely examples of some of the best examples of such case studies, especially those employing various kinds of mixed methods. Donors should also consider conducting a study or meta-analysis of independent case study evaluations to learn more about the approaches employed and the quality of the products. While independent evaluations have increased in number, there is little understanding of what is being learned from these evaluations in the development community but also what more could, and should, be learned if more rigorous or innovative approaches were promoted. As it stands now, vastly more significant resources (from a learning and innovation standpoint) appear to be devoted to 'pushing the envelope' with experimental or quasi-experimental approaches (e.g., very little in the way of resources and publicity have been devoted to making best practices in case study methods more rigorous and accessible by USAID's Learning Team in the Center for Rights, Democracy, and Governance). At the same time, far too few independent evaluations are commissioned to look retrospectively at project interventions over a longer period of time to understand the incremental, but often more impressive cumulative influences that development projects have on institutional reform and change leadership years after the projects have ended.

The fact is that case studies are often taken seriously as a way to communicate important information and learning about the influence of, and progress made by, development programs. Indeed, a widely cited critique of American democracy programs—advanced by Carothers—is based on a set of four cases.[54] As we have discussed, the advantage of case studies has been in the scope of things they could

---

Nakamura and Samir Musovic, Success Where You Least Expect It: How Parliamentary Oversight Produced Better Government in Bosnia Herzegovina, Paper delivered at the 2012 Annual Meeting of the International Political Science Association, Madrid.

[52] Greg Power, The Logframe and the Beautiful Game, and Enabling Change: A Behavioral Approach to Political Programming, Agile Programming Paper 2, Global Partners. .

[53] Robert Nakamura, Heather Senecal, and Andrea Wolfe, op.cit.

[54] Thomas Carothers, 1999. *Aiding Democracy Abroad: The Learning Curve* (Washington: Carnegie Endowment for International Peace) The four cases are derived from democracy assistance programs in Guatemala, Nepal, Zambia, and Romania. A more explicitly comparative approach is taken by Carothers in "The Reagan Years: The 1980s" in which he looks at four different Reagan administration policies with promotion of democracy as their stated goal in Latin America. He concludes that

consider and in the flexibility they offer for appreciating how variables operate in the context of actual circumstances. The weakness of the approach of course also lies in its capacity to appreciate the unique. As a result, the evaluation literacy level needs to be raised so that these strengths and weaknesses can be better appreciated and particularly mixed methods used where applicable to focus attention on particular concerns. One method that has been consistently underutilized is direct observation—something that could play a bigger role in legislative development work, particularly as regards changes in MP behaviors carrying out representative functions their home districts and in working with staff; how staff work routines have changed, etc.

Again, since the only evidence for conclusions is drawn from the case or cases studied, the applicability of their conclusions can be challenged for relying on too narrow a base of evidence. So not only should these conclusions in single cases be qualified as appropriate due to the particular evidence base, but generalizations from case studies should always require a rationale explaining why they can or cannot be applied more widely. In addition, there are numerous approaches to conducting case studies and their variety makes it difficult to compare results and to aggregate experiences. That being said, these various approaches should be more widely discussed, circulated, and compared, with the goal of promoting a new generation of independent evaluations focused on the most usable knowledge for practitioners and the development community as a whole.

Comparative Case Study and Field Network Methods:

As the above examples indicate, generalizations from a base of several case studies have achieved *de facto* acceptance with caveats in the democracy evaluation area. It is, however, possible to press the case study method further and provide a firmer foundation for drawing broader conclusions with more confidence.

An important methodological problem posed by case study methods is expressed in the question "compared to what?" Another problem is encountered as cases from different sources are combined for analytic purposes and the diversity of their methods makes aggregation problematic. We will address both these issues next. We propose using (1) a research design based on a comparative case approach to answer the "compared to what" question; and (2) incorporation of a field network method for this and subsequent studies to provide a broader and firmer basis over time for generalizations.

We note, before proceeding, that the case study approach we are recommending should be considered as an addition rather than as a replacement for individual case studies as they are currently conducted in all their diversity – although again, donors and practitioners should both be seeking to enhance the rigor of case studies within this otherwise diverse universe. That is, as in most areas where the relationships between variables are fluid, not well understood, and definitions of success are contested, a reliance on a multiplicity of methods and indicators useful for different things is often considered the best approach.

---

democracy was less the goal than other goals. In Abraham Lowenthal, *et al. ed.,*1991.*Exporting Democracy* (Baltimore: Johns Hopkins Press).

1. Comparative Case Method.

We want to gather data to draw lessons about what has worked and under what circumstances. This approach has been employed explicitly in the evaluation of American environmental programs[55] and implicitly in the assessment of democratization efforts in Central America, to point out just two examples.[56]

How does this approach work? Assume that the universe of cases consists of countries in which USAID is trying to achieve the same or similar goals such as legislative and democratic functionality by using a relatively small number of different strategies/approaches. Our first problem is to identify the sets of countries and approaches we are most interested in learning about. Second, we have to identify specific countries that are most representative of type and of exemplary implementations of particular approaches.

The countries thus selected are treated as if they were sets of "natural experiments" in which important determinants were varied by circumstances to produce different results. So the rationale for generalizing from them is similar to that presented by laboratory experiments in which variation in treatment results from experimenter choice rather than the world.[57] Here again, however, sufficient resources need to be made available to generate a sufficient number of sets of comparative country and approach variables. All too often, even within resource-constrained operating environments, donors have failed to direct adequate funding to larger-scale and longer-term comparative case study efforts that would seek to synthesize some of the truly most valuable learning from multiple institutional reform projects.

2. Field Network Method:

The second issue is that of building a base of case studies that are consistent enough in approach and method to support aggregation. In the short run, if the same investigator or team conducts the initial set of studies, consistency is less of a concern, although an explicit master design should be used and adapted to circumstances. In the long run, however, a more explicit plan for building comparability should be implemented.

The field network approach has proven useful as a means for creating a large base of case studies that can be usefully compared and serve as the basis for persuasive generalizations. The technique has been used for over a decade in the evaluation of changes in welfare policy whose circumstances have some important similarities with legislative development efforts: for example, the same general goals apply to all cases, diverse implementation strategies have been employed, and substantial variation in local circumstances is always present.[58]

---

[55] This approach has been used in the evaluation of Superfund, see Church and Nakamura, Cleaning Up the Mess (Washington: Brookings, 1993).

[56] See Thomas Carothers, In the Name of Democracy: US Policy Toward Latin America in the Reagan Years, University of California Press, 1991..

[57] See Donald Campbell, "Reforms as Experiments."

[58] This approach was originally devised by Richard Nathan as a means for understanding, in depth and with appreciation of local circumstances, in national programs implemented through a federal system.

The Rockefeller Institute coordinated welfare evaluation work in over 20 states and 50 specific sites using the field network approach.[59] Richard Nathan managed other Rockefeller Institute studies of revenue sharing and education (in the case of the "No Child Left Behind") using the same approach. Each drew on the work of associated scholars in the different states, working from an explicit set of topics for investigation. Coordination was achieved through careful monitoring by the Institute's central team and frequent meetings and other forms of exchange.

While the Rockefeller Institute studies used the field network to capture what is happening in diverse places at the same time, many of its features can be incorporated into a means for carrying out studies of developments in many countries over a longer period of time. What is achieved through meetings and interaction among researchers, for example, could be provided by training, more explicit protocols, and a more conscious process of emersion in a developing body of case studies incrementally assembled by this technique. The organizational permanence of USAID or DFID could be an important asset in producing the basis for this enterprise.

Conclusions

"A way of seeing is a way of not seeing."[60] We started with a summary of how social science work (in psychology, political science, and public management) in evaluation has been adapted to the needs of donor agencies (for credit and control), and through them shaped the conceptual framework through which the implementation of legislative development programs are run and evaluated. This way sees legislative development from a largely top down perspective, evaluating programs by fidelity to pre-determined goals and approaches, and producing a framework for managing efforts to prevent diversion of resources. While useful for "seeing" many things (clarifying goals, program logic, responsibilities, safeguards for resources), this framework presents a number of problems, including (1) mismatches when scholarly concerns are "projectivized" to achieve bureaucratic purposes, and (2) in the privileging of sometimes inappropriate conceptions of methodological rigor. It is also a way of "not seeing" the ground level problems of implementing legislative programs in environments that are fluid, politicized, multi-faceted, and within which goals change with internal and external circumstances. An emerging consensus of interested scholars and practitioners – one that may finally be reaching a critical mass – sees that achieving change in the desired direction requires the ability to adapt to circumstances and thus modify pre-determined policies and re-program resources. This parallels the lessons of the first and second generation of implementation scholars on the importance of mutual adaptation, which is still part of the contemporary public management curriculum.

Yet if problems continue to persist (and perspectives are distorted in other ways) through another kind of narrowing of vision through the superimposition of inappropriate and often very costly experimental methods on top of a dominant evaluation method), the solution is less one of wholesale criticism or replacement of the existing standardized logframe-driven approach than it is of introducing other perspectives to cover the blind spots. We suggest supplemental ways of "seeing" through the use of

[59] See, for example, the literature on welfare reform following the adoption of the Personal Responsibility and Work Opportunity Act of 1996. The Rockefeller Institute using a field network approach is running a large set of such studies encompassing 20 states and 50 local programs.

[60] Kenneth Burke, The Philosophy of the Literary Form, University of California Press, 1941.

existing case study methods already used by scholars to evaluate policies in comparable areas with similar problems (uncertain technologies, contested goals, diverse environments). We also endorse the use of a variety of process-oriented evaluation methods to better identify the often very significant, though ostensibly modest, influences that institutional development projects have on development counterparts in terms of their behaviors and relationships. Finally, we strongly recommend that donors elevate the importance of these topics in development project evaluation discourse and actively encourage more rigorous and innovative use of these tools in the case study universe, deliberately seeking to learn from different case study approaches, including comparative case studies and field network methodologies. All of these approaches are useful for drawing design and implementation lessons, because they focus on real situations, sequences, problems, and look directly at the places where expectations about change meet existing practices.