

## **APPENDIX 3: A PRACTICAL EVALUATION PROCESS<sup>1</sup>**

### **I. Introduction**

1. This appendix sets forth a Practical Evaluation Process (PEP) for assessing the kinds of legal empowerment impact previously described in this report. It responds to the interest on the part of development organizations in quantitative indicators that can measure the results of legal empowerment. As an alternative to the indicator-driven approach that has proven to have a mix of strengths and weaknesses in the closely related field of democracy assistance, a different perspective and associated mechanisms are proposed for application by development organizations.

2. Although this is referred to as a Practical Evaluation Process, it pertains to monitoring project performance as well. The distinction is that monitoring is supposed to inform an organization as to whether it is achieving its desired results. In theory, monitoring does not explain why it may be achieving or falling short of its targets. To get at the "why's," development agencies can conduct evaluations. In the case of legal empowerment, however, the two processes converge to such a degree that both are addressed under the rubric of PEP.

3. As a starting point, what is the rationale for thinking in terms other than indicators? Some persons within and outside institutions that employ indicators for assessing democratic assistance have found the approach counterproductive. The reasons for this include the rigidity of the indicators approach; its reliance on numbers that may be ill-suited to programs; the failure to establish causal links between apparent progress and the impact of donor investments; and a variety of other factors.<sup>2</sup> For example, as Carothers concludes, "The indicator system not only fails as a method for assessing the effects of democracy aid programs, it introduces distortions into the aid itself."<sup>3</sup> He goes on to assert that "the indicator system is extremely unpopular among USAID [United States Agency for International Development] officials, at least those working on democracy building."<sup>4</sup>

4. Indicators should not be totally dismissed as irrelevant to legal empowerment. Rather, they should be placed in a context where they may lend themselves to certain programs and to measuring certain kinds of progress where appropriate. Since use of indicators is not always appropriate, PEP reflects the view that they should not dominate evaluation of legal empowerment or related work.

5. Guided by the above considerations, this discussion of PEP addresses legal empowerment evaluation mechanisms, the incorporation of evaluation planning into program design, and evaluation reporting.

---

<sup>1</sup> Appendix 3 substantially draws on research and related work conducted by LES co-team leader and Overview Report co-author Stephen Golub under an Open Society Institute Project Fellowship. An expanded and modified version of the appendix will separately be published as a product of that fellowship.

<sup>2</sup> For a partial accounting of these drawbacks, see Thomas Carothers, *Aiding Democracy Abroad: The Learning Curve*, (Washington: Carnegie Endowment for International Peace), 1999, p. 281-302.

<sup>3</sup> Carothers, p. 294.

<sup>4</sup> Carothers, p. 296.

## II. Evaluation Mechanisms

6. This section describes a range of relatively basic mechanisms that the Asian Development Bank (ADB) and other international organizations may consider using in evaluating legal empowerment programs. The mechanisms pertain to the types of impact previously detailed in the legal empowerment report. While the menu is far from exhaustive, it does seek to capture some of the practical ways in which development institutions gather information. It further aims to demonstrate how some of these approaches can be employed informally in connection with legal empowerment evaluation.

7. The degree to which an institution uses any of these mechanisms must hinge on the resources available under a given program. Clearly, small-scale programs should not generate large-scale evaluations. Which mechanisms to use and which questions to ask also depend on the nature of the program.

8. Evaluation mechanisms serve to assess at least two levels of impact. The first is *program-specific* or *direct* impact on organizations and individuals directly funded, trained, or otherwise assisted by a given project. *Indirect impact*, in contrast, results from a ripple effect through which other groups and persons are in turn affected by those directly impacted. It also can apply to effects on broader populations, segments of a society, or the society as a whole—that is, as long as a contributory role can be established for the organization in question.

9. Certain evaluation mechanisms also can serve two other purposes. First, they can provide *general insights* about a society or issue, information that can help shape or guide a development agency's work even if it does not ultimately aid an evaluation. Second, certain tools can help an agency and its development partners assess the *technical competence* with which an activity has been carried out (for example, participants' impressions of the speakers, materials, and venue of a conference). This can be the case even if they do not reveal the impact of the activity (such as what participants learned at the conference, and what use they eventually made of that knowledge).

10. Evaluation mechanisms, and the information they provide, can be divided into two categories: *quantitative* and *qualitative*. Very basically, the quantitative mechanisms gather information that can be best expressed numerically, such as the number of people trained or the percentage of a population that understands certain rights.

11. Qualitative mechanisms, in contrast, gather data that is best expressed and recorded in a non-numerical manner, such as individuals' detailed opinions, experiences, or observations regarding programs, or their reports on how they subsequently made use of the knowledge or skills acquired. They also may better assess a project's unanticipated impact.

12. As both evaluation mechanisms must vary widely among programs, it is counterproductive to dictate a strict formula for employing them. The following partial menu of mechanism is but a starting point for developing an evaluation. Which to select hinges on the nature of the program.

### Sample Surveys of Program Participants/Beneficiaries

13. The most straightforward way of assessing program-specific impact on knowledge is a series of oral or written surveys. International organizations can determine whether a legal empowerment initiative has increased participants' knowledge by surveying a sample of them at

the outset of the initiative, and at least two additional times—at the close of the program and at one or more later dates.

14. If appropriate, the survey at the close of the program should include questions that focus on the technical competence of its implementation, as well as questions that measure impact. Assessing competence can be useful for determining whether the approach to similar activities needs to be refined in the future. It should be clear that the answers of respondents regarding technical competence say little about the impact of the approach. The respondents may, for example, report that the speakers were excellent and that they learned a great deal; but these responses do not reveal what they actually learned.

15. In addition to knowledge, impact-oriented surveys at the outset and close of a program can ascertain changes in attitudes. They can also determine improvements in skill levels, where asking questions (for example, about how to prepare an affidavit or obtain information from a government bureau) demonstrates whether program participants retain at least a theoretical knowledge of how to exercise legal skills.

16. The follow-up survey, undertaken at least six months later, provides a much better gauge of whether the participants have retained knowledge, skills, or attitudinal changes initially generated by the program. As such, it yields a much more significant reflection of impact than does the survey undertaken at the program's close.

17. Surveys can generate both quantitative and qualitative data. The former is best gathered through yes/no or multiple choice questions. It can be especially useful if both pre-program and follow-up surveys are possible.

18. By employing open-ended questions, this mechanism also can generate important qualitative information. The follow-up survey can solicit information about behavioral changes, organizational improvements and government actions to which the program might have contributed. Though some responses to questions of this kind can be quantified, the real usefulness (as described below) is that they could provide the basis for subsequent qualitative inquiries that verify and expand understanding of the impact.

19. Of course, many programs may not lend themselves to surveys at each of the stages described above. A program may aim for more general diffusion of knowledge than can be measured by a set list of questions at its outset. The fact that surveys can help document diverse types of impact accordingly suggests that follow-up surveys be conducted even if their questions differ from those originally asked of participants (because the program's perceived impact has changed), or if no initial surveys were conducted at all. The data need not be comparative for it to be illuminating.

20. How the information is initially and subsequently collected depends substantially on such factors as the resources and communications facilities available, the nature of the participants, and the project into which a legal empowerment initiative might be integrated. Under some circumstances, for example, house-to-house interviews might be necessary. Under others, questionnaires could be distributed and collected at meetings organized by an NGO or government agency.

### **Sample Surveys of Broader Populations**

21. Under circumstances where an institution anticipates indirect impact (that is, impact reaching beyond program participants), it should consider surveying populations likely to be affected. They need not be as broad as the general population. Rather, they could be residents of communities where a development organization undertakes activities. Ideally, such polling should be conducted before the program begins, and again as a follow-up at least six months later.

22. In conducting surveys of this kind, the organization should ask respondents not just *what* they know (or believe, or do), but *how* they came to know it. Respondents may not be able to identify whose training they attended, whose pamphlet they received, or whose radio show they heard. Their answers nevertheless may provide bases for concluding whether a given effort contributed to their enhanced knowledge.

23. In addition to revealing impact on citizens' attributes or circumstances, polling a broader population can suggest or verify other types of impact. For example, it can indicate whether a given NGO's or government department's capacities have improved.

### **Control Group Surveys**

24. Whenever possible and appropriate, an organization could poll a control group that is not participating in or affected by its training or other activities. Where surveys or other evaluation tools suggest that the organization has contributed to impact, a control group's unchanged attributes will tend to verify the program's success. Preferably, the control group that is surveyed initially should be the same as that surveyed at the follow-up stage. Again, practical considerations may impinge on this. If so, a new control group still would offer a basis for comparison.

### **Baseline Surveys**

25. Where funding and circumstances permit, baseline surveys conducted by an organization when it first launches a program can determine the level of relevant knowledge and other attributes of the general population or significant segments of it. While such surveys may primarily be conducted to generate general insights that can shape a program, they can also complement or substitute for initial surveys of control groups.

### **Qualitative Interviews and Discussions with Participants**

26. Qualitative interviews and discussions with participants seek information that cannot be translated into survey numbers. The two can go hand in hand, with an oral survey turning into an open-ended discussion. Of even greater importance, interesting survey responses can trigger a subsequent interview.

27. Such interviews can take many forms, but the two most prominent might be called "semi-structured" and "open-ended." As understood here, the former starts with a common set of questions asked of all participants interviewed, before branching out to cover whatever interesting topics arise in the course of discussion. The latter focuses on whatever issues seem most pertinent with respect to a given participant.

28. As with the surveys, the most useful interviews could well take place months or even years after a program has ended. It is at this stage that the respondents might be best able to analyze what use they or others made of the program.

### **Qualitative Verification Through Third-Party Interviews**

29. The most important element of evaluation is independent verification of impact. Wherever practical, an organization should consult knowledgeable third parties to seek such verification. Although it is useful for the organization itself to perceive positive impact regarding such factors as behavior, government responsiveness, or material circumstances, it is highly desirable to confirm that information in whole or in part. Independent sources of confirmation can include journalists, academics, government officials, NGO leaders, and representatives of international development organizations.

30. Under some circumstances this relatively in-depth evaluation of relatively modest activities would be too expensive, time-consuming, or politically sensitive. It also would be unnecessary where the only goal of the activities is to affect participants' knowledge or attitudes, and these can be determined through quantitative surveys. At the same time, independent verification would lend great credibility to the organization and its partners alike.

### **Verification Through Textual Analysis**

31. Where a legal empowerment initiative makes a significant contribution to government policies, it can and should confirm this through textual analysis: comparing policy proposals with the resulting regulations, legislation, or other documents. Even in cases where an initiative affects just one aspect of a government policy, textual analysis comparing its suggestions with the revised policy can demonstrate impact.

### **Case Studies**

32. While legal empowerment assistance relates to the core dynamics of how a community operates, it often lacks in-depth insight into the dynamics of change that it helps bring about. Case studies—that is, focused research on actual or potential instances of impact—represent an effort to achieve and learn from such understanding.

33. As the term is employed here, a case study consists of an in-depth look at impact as diverse as how NGOs successfully pushed adoption of a piece of legislation or how a community group helped to reduce the incidence of violence against women. It can also document partial successes or failures, yielding useful lessons. In the former instance, such scrutiny can consist of a social scientist tracking a legislative bill's progress as it occurs, or retrospectively learning as much as she can about the strategies, politics, and other influences that affected its form and course. In the latter example, it could involve a quasi-anthropological observation of one or more communities over time, to learn what strategies and forces helped (or failed) to overcome deeply ingrained attitudes and practices.

34. As is evident, a case study can draw on social science research tools, including some of the other evaluation mechanisms described above. More than any other mechanism, it offers opportunities to learn about strategies for effecting change at both national and grassroots levels.

35. An obvious difficulty with this approach is that it can be most useful when a bill or community is tracked from the start of a legal empowerment intervention, but that intervention may not produce success or even useful lessons. In some cases this can be overcome by researching a number of bills or communities. In other instances, it may prove worthwhile to review progress retrospectively. Regardless of the approach taken, the time and expense involved with such studies preclude their being appropriate for certain activities. In planning a study of this kind, cost considerations should be borne in mind. For example, in many countries a local academic specialist and graduate students can be contracted for several months for less than the cost of engaging a Western consultant for a month. The former typically have greater insight into their societies.

### **III. Incorporating Evaluation Planning into Program Design**

36. Is it worth the effort to plan an evaluation strategy at the outset of a legal empowerment initiative? Not always. The duration of assistance and the unanticipated activities and impact that flow from them may minimize the value of planning ahead in this manner.

37. Nevertheless, there also are many instances where the duration of assistance and the nature of the work can be sufficiently determined that it makes sense to construct an evaluation plan well in advance. All elements of the plan should be *illustrative* in nature, however, rather than constituting absolute commitments regarding what an organization will achieve and how it will evaluate impact. Even with a known quantity, many factors relating to the program can change before it starts or as it proceeds.

38. The elements of the plan need to be suited to the project. Examples of possible elements include:

#### **Anticipated Impact**

39. The plan could state the nature of the type(s) of impact an organization aims to achieve (for example, regarding citizen knowledge, government responsiveness, or other factors). It also could state the anticipated scale of such impact, if possible. This applies to such matters as the number of citizens, communities, or policies affected.

40. The anticipated impact is illustrative of what an organization hopes to achieve. It is *one* potential basis for evaluating a program. It may in fact turn out to be the sole basis for evaluation. Where other types of impact materialize, anticipated impact may become less relevant, or even irrelevant.

#### **Unanticipated Impact**

41. It is important to try to predict at the outset what a program will accomplish, to set standards by which it can be assessed. Yet an exclusive focus on and investment in such prediction can delay or derail a legal empowerment initiative to no good end, when the reality of unintended consequences catches up with it. For example, people who learn about health benefit rights through a community health project might become engaged in good governance activities that have nothing to do with those specific rights. Farmers who receive training in land tenure laws may strengthen their previously weak community-based organization. They may even provide input into regulatory reform regarding land rights. None of these results can necessarily be anticipated.

42. With this in mind, an initiative must be open to seeking out whatever impact, good or bad, flows from its activities. The best results may not be planned. Even if there are counterproductive consequences, these need to be determined and understood in order to improve a project and avoid mistakes in the future.

### **Planning of Evaluation Mechanisms**

43. The choice of evaluation mechanisms hinges on the nature of the program's anticipated impact. Determining citizen knowledge sometimes lends itself to surveys; assessing their input into law or regulatory reform is much more qualitative.

44. It also depends on the program's duration and budget. A few third-party interviews may well make sense for a relatively modest project. In-depth case studies may only be justified for a more major effort.

45. Like impact, evaluation mechanisms can only be "anticipated" in the evaluation plan. They may require modification or replacement once the program starts. This may be due to unexpected changes relating to the program, its apparent impact, or the broader society.

46. In selecting such mechanisms, the plan could also state illustrative questions that the organization will ask in assessing impact. Again, however, it may need to modify these questions by the time the program begins or as it proceeds.

47. In addition to mechanisms for assessing anticipated impact, the evaluation plan should explicitly include at least one method of identifying unanticipated impact. The two can be one and the same. Where, for example, follow-up interviews aim to determine anticipated impact, they also can probe for benefits that were considered possible but not likely at the outset of the program, or that were not considered at all. A survey that mainly probes for quantitative data of a predetermined nature may also include open-ended questions for the respondents.

### **Time Frame**

48. The plan could state a timetable for utilizing the anticipated evaluation mechanisms. Because so much of the impact of legal empowerment assistance may only materialize long after the completion of a program, a final evaluation should ideally take place at least six months after the program's completion.

49. For a major program, an organization may want to conduct more than one post-program evaluation activity, doing the final one a few years after program completion. Obviously this cannot be done for most programs, but occasional use of this device is worthy investment. This may well be the most useful means of determining whether a program ultimately did any good, and the best way of drawing lessons from the experience. Given the realities of project funding, some organizations could cover the evaluation activity out of an indirect cost pool or funding categories that are not project-specific. Since evaluation benefits an organization as a whole (rather than just a specific program), standard accounting norms should ideally justify this expenditure.

### **Resources**

50. Finally, an evaluation plan should estimate—and the program budget should allocate—the resources an organization needs to carry out evaluation. As with all other elements of the

plan, the budget estimate should be illustrative. That is, an organization should set aside a certain amount of funds for evaluation from the outset, but should employ those funds flexibly as it becomes clearer which information and mechanisms will best contribute to useful evaluation.

#### **IV. Evaluation Reporting**

51. How can a development organization integrate and weigh the information generated by its evaluation mechanisms? What types of reports suit these purposes? What kinds of questions could the reports address? Finally, but most fundamentally, how can the organization use and learn from the reports? As with other aspects of PEP, the answers hinge on the nature of the organization and project. This section nevertheless hazards some suggestions.

#### **Weighing Impact Priorities in Reports**

52. Relatively modest and short-term legal empowerment programs are most appropriately evaluated in terms of impact on citizen knowledge and attitudes. The more ambitious and long-term a program, the more it should be assessed in terms of citizen behavior, material circumstances, good governance, poverty reduction, and law/regulatory reform and implementation.

53. Why these broad distinctions? Knowledge often represents a first step toward changes in behavior. Changes in material circumstances are important if they are pursued as ends in themselves, as with farmers organizing and learning about the law in order to realize the benefits of agrarian reform. Yet they are equally important if one takes a purely process-oriented view of them, for they can serve as the firmest proof of behavior, participation, or other types of impact.

54. The same dual role applies to impact regarding law reform, legal implementation, and other government actions. These may be sought in and of themselves. Or, for those taking a process-oriented perspective, success regarding these matters may be seen as proof of impact regarding citizen participation.

#### **Questions for Evaluation Reports**

55. Unless the answers are self-evident, an evaluation report should consider addressing as many of the following questions as possible, whether explicitly or implicitly:

56. *What is the nature of the impact?* It is not always essential to break down the discussion into distinct analyses of knowledge, behavior, good governance, or other types of impact. While the nature of impact is rarely so neatly compartmentalized, an organization should keep the various kinds of impact in mind in preparing the report.

57. *What is the scale of the impact?* It is not always possible or appropriate, but an organization can try to state the number of people or communities affected, or the population actually or potentially affected by government actions to which a program has contributed.

58. *What is the significance of the impact?* This element addresses the crucial issue of context. In some instances, the significance of a particular impact may be self-explanatory. In others, it may be necessary to explain why cultural, political, economic, or other factors in the host society make an achievement that might seem minor in one country a milestone in another. Or it may be advisable to illuminate the importance of a particular policy achievement, or the

ripple effect of impact on a specific community. Conversely, the report might need to provide context that addresses why the project fell short of expectations or its potential.

59. *What is the legal empowerment program's contribution to the impact?* An organization should explain how it helped bring about the impact. The emphasis, of course, is on *contribution* to impact, rather than sole credit for changes that came about.

60. *What is the significance of the organization's contribution?* The importance of an organization's role may be self-explanatory, or may require some further illumination—as, for example, what might have happened (or not happened) had its program not taken place.

61. *What evaluation mechanisms were employed, and why?* This question goes to the heart of verification, and how it can be determined that impact took place and that the organization contributed to it. It also can address how the organization reached certain conclusions. This becomes important if it cannot clearly document these matters. Particularly for outside audiences, it could be helpful to explain why the evaluation report documents them as much as is reasonably possible, while acknowledging possible gaps in what can really be known about impact and attribution.

62. *What lessons have been learned?* This can be the most important element in the evaluation. The lessons can range from the relatively mundane (specific techniques for conducting educational seminars, for example) to strategic (such as considerations regarding whether and how to work with a government agency permeated by corruption).

63. *Who will see the evaluation?* This is a question of a different nature, but it is still worth asking—as well as answering in favor of dissemination to a variety of interested individuals and organizations. Too many evaluations end up collecting dust on the shelves of the organization that commissioned them, while other organizations scramble to reinvent the wheel, unaware of valuable lessons learned by previous or similar efforts. This situation contributes to duplication of effort, and failure to learn from past experience.