How Much Does Evaluation Matter?

Some Examples of the Utilization of the Evaluation of the World Bank's Anti-Corruption Activities

MITA MARRA

Operations Evaluation Department of the World Bank, USA

This article offers empirical evidence of the utilization of evaluation findings of the World Bank Institute's (WBI) efforts to help reduce corruption in Tanzania and Uganda. These initiatives are part of the World Bank-WBI program to curb corruption in developing countries. This analysis focuses on the mid-term evaluation of the WBI's anti-corruption activities in those countries. The study shows, through a series of examples, how evaluation has been used in both an instrumental and an enlightenment fashion by program designers and implementers. Although links between knowledge generation and utilization are seldom clear and direct, and specific information cannot always be isolated as the basis for a particular decision, the examples show that utilization has occurred, bringing about change in program design and implementation.

Generalizations from evaluation can percolate into the stock of knowledge that participants draw on. Empirical research has confirmed this. . . . Decision makers indicate a strong belief that they are influenced by the ideas and arguments that have their origins in research and evaluation. . . . The phenomenon has come to be known as 'enlightenment'. . . . (Weiss, 1998a: 25)

Introduction¹

This article offers empirical evidence of the utilization of evaluation findings of the World Bank Institute's (WBI)² efforts to help reduce corruption in Tanzania and Uganda. The initiatives are part of the World Bank-WBI program to curb corruption in developing countries. This analysis takes into consideration the mid-term evaluation of the WBI's anti-corruption activities in those countries, entitled 'EDI's Anti-Corruption Initiatives in Uganda and Tanzania' (Leeuw et al., 1998).³ The study is part of the World Bank Institute's Evaluation Studies' (WBIES) Evaluation Utilization series, which aims to understand how and under what circumstances evaluation studies are utilized.⁴ This article focuses on the

task managers' view of how evaluation-informed knowledge has resulted in modifying program implementation and goals. The methodology adopted consists of semi-structured interviews conducted with the task managers and the principal evaluator, a content analysis of the evaluation report and an extensive review of the literature on research and evaluation utilization and organizational learning.

The analysis uses two explanatory models: the *instrumental* model and the *enlightenment* model. The instrumental perspective assumes a rational decision-making process: decision makers have clear goals, seek direct attainment of these goals and have access to relevant information. From the enlightenment perspective, on the other hand, users base their decisions on a gradual accumulation and synthesis of information. Weiss describes enlightenment evaluation as providing a background of working understandings and ideas that 'creep' into decision deliberations (see Rossman and Rallis, 1998). Evaluation is used not only in action but also in thinking, and results are incorporated gradually into the user's overall frame of reference (Weiss, 1997). As Vedung (1998) notes, program implementers, designers and stakeholders receive cognitive and normative insights through the evaluation, which enables them to thoroughly scrutinize the premises of the program and gain a deeper understanding of its merits and limitations.

The two models are complementary rather than mutually exclusive. Both are construed as empirical theories. As shown in the following paragraphs, there are cases in which policymakers set clear targets and, by means of evaluation, succeed in achieving them in ways that resemble instrumental use. In most situations, however, real life deviates from the instrumental model. As Vedung (1998) says, 'Evaluations may have significant consequences but not in the linear sequence asserted by the instrumental view. The role of evaluation in political and administrative contexts is many sided, subtle and complex'.

The Content of the Program

As part of its assistance to client countries to help control and curb corruption, the WBI has developed the concept of national integrity systems as a means to identify and strengthen those institutions with a mandate to fight corruption.

Three activities are at the core of the anti-corruption approach:

- integrity workshops;
- media workshops;
- service delivery surveys.

At the request of national governments, the WBI helps to develop and institutionalize an anti-corruption program that will be supported by the public, the government and civil society. It does so mainly by organizing integrity workshops at the national or regional level or for specific sectors such as the media, and by helping to conduct service delivery surveys (SDS). The main purpose of these integrity workshops is to formulate and agree on an anti-corruption program, and in the process raise awareness of the costs of corruption and discuss the roles that various institutions – *pillars of integrity* – can play in the fight against corruption.

The workshops also serve as forums for stimulating policy dialogue among the integrity pillar institutions, with the goal of developing an outline of a national integrity system geared to curbing corruption.

Within this broad framework of common intention and understanding, the media workshops are the key players in informing the public about corruption and exposing corrupt practices. The first generation of media workshops – investigative journalism – focused on the media's role in raising awareness of corruption and on improving investigative techniques. Journalists were given basic training in the skills needed to carry out investigations. They learned to obtain information in ways that are ethical and respect privacy, and to avoid litigation. This training was delivered through (1) mini press conferences; (2) simulation exercises called 'Freedonia';⁵ and (3) field trips. There is now a second-generation course on advanced investigative journalism and an investigative journalism workshop for editors.

The service delivery surveys, undertaken in partnership with CIET International, are measurement tools that combine social and economic data with information on citizens' experience, expectations and perceptions of service delivery. This is done through a combination of techniques – analysis of available data, household surveys, focus group discussions, key informant interviews, institutional reviews and observational studies. To build local capacity to carry out this kind of analysis, local counterparts are trained and participate in all aspects of this process.

Intended Use of the Evaluation

The mid-term evaluation of the WBI's anti-corruption activities in Tanzania and Uganda was commissioned in 1997 to shed light on the strengths, weaknesses and impacts of the activities as they had unfolded. The evaluation was requested by WBI management in view of the pending expansion of the program to at least 15 other countries, especially in Central and Eastern Europe, where task managers will need to address corruption in a region with totally different historical, political, economic and social characteristics. Weiss (1998a) makes the point that this is the kind of knowledge that can enrich thinking and enable those responsible to proceed with greater conviction, often in difficult circumstances.

The evaluation highlighted two kinds of concerns:

- Resource management and quality assurance within a large-scale program.
- The suitability of the program's conceptual framework and design for a different context and audience.

First and foremost, institutionalizing the program implies that financial and human resource allocation decisions are to be taken. Hirschman (1967) observes that the expansion may be accompanied by a qualitative deterioration of performance and results, not only because of the need for more inputs, but because of lack of experience and knowledge of how to carry out program activities in different contexts.

The evaluation therefore focused on program costs and management in order to

determine (1) which procedures and techniques were more or less successful; (2) which were achieving results most efficiently and economically; (3) which features of the program were essential; and (4) which could be changed or dropped (Weiss, 1998a). Such retrospective data are relevant, as Rist argues (1999), in order to (i) learn how the organizational response to a problematic situation is conceptualized; (ii) what expertise and interest are shown by management and staff; (iii) what controls over the allocation of resources are in place; (iv) whether the organizational structure reflects organizational goals; (v) what means exist to decide between competing demands; and (vi) what kinds of interactive information or feedback loops are in place to assist managers in their ongoing efforts to move the program toward stated objectives. This type of information on the implementation process is critical to managers as they struggle to move toward organizational goals.

The purpose of the mid-term evaluation, therefore, was to gain both cognitive and empirical insights that would feed future decisions and actions and at the same time, trigger an organizational learning process. Argyris and Schon (1978) argue that evaluation provides a means of formulating collective lessons of experience and incorporating them into organizational theories of action.⁶ Individuals within organizations reflect critically on their own behavior, identify the ways they often inadvertently contribute to the organization's problems, and then change how they act. Evaluation thus plays a crucial role in promoting what Argyris calls 'double-loop organizational learning' (Argyris, 1994).⁷

Furthermore, larger-scope program activities and new recipients usually require *ad hoc* adjustments and specific targeting. This raises structural issues regarding the program design and conceptual framework. Africa is substantially different from Central and Eastern Europe, and what works in Tanzania and Uganda might be not appropriate in Russia. Thus, the theoretical assumptions upon which the program was built might no longer hold, particularly since all social programs wrestle with prevailing contextual conditions.

'Programs are always introduced into pre-existing social contexts and these prevailing social conditions are of crucial importance' (Pawson and Tilley, 1997). These two authors refer to the 'embeddedness of all human action within a wide range of social processes' as the 'stratified nature of social reality'. 'This vision of a stratified reality', they argue, 'leads directly to the notion of the *explanatory mechanism*. This is a useful metaphor, since it captures the idea that things work by going beneath their surface (observable) appearance and delving into their inner (hidden) workings'.

A crucial task of evaluation is to investigate the extent to which pre-existing social contexts 'enable' or 'disable' the intended mechanisms of change. A global assessment of the pilot program's functioning was, therefore, expected to help redefine its inherent characteristics so they can be adjusted to a diverse context.

In Search of Users: Implications for Dissemination

When the evaluation was commissioned, it was impossible for the evaluation team to identify all potential users. Yet evaluators were able to target some specific intended users, inside and outside the World Bank.

The metaphor of concentric circles is a useful visual way of depicting how evaluators identified their stakeholders. They began with the task managers directly involved in the design and management of anti-corruption activities; then moved to the WBI at large, which has an explicit mandate to combat corruption; and then to the Bank's operational sectors, who are involved in lending operations and consulting with governments in developing countries; to client governments that benefited from anti-corruption activities; and finally to civil society as a whole. Writes Weiss:

The reason for adding civil society to the user category is that evaluations of any moment involve issues that go beyond the interests of the people involved in the kinds of programs under study. Broader publics need to hear about the successes and shortfalls of program implementation and program outcomes. Active in many local initiatives, members of civil society use evaluative information in the program activities in which they are engaged as volunteers, board members, and advisors. More than that, they are opinion leaders in their communities. They can use evaluation to illustrate the successes that programs can have and help to counteract the apathy and hostility that many social programs face these days. (Weiss, 1998b: 29)

In sum, evaluators targeted intended users mainly within the World Bank with evaluative questions (see below) building upon these users' information needs. The evaluators' intent, however, was to reach as many potential users as possible by raising awareness of their findings and making specific recommendations for anti-corruption actions. Their identification of different types of users mirrors the different ways that evaluation can be utilized.

The previously mentioned concepts of instrumental and enlightenment models at this point help to clarify two different ways of evaluation utilization. Instrumental means that evaluation findings are utilized as a means in goal-directed problem-solving processes – that is, used *strictu sensu* – whereas enlightenment suggests that detailed findings become generalizations that are eventually accepted as truths and come to shape the ways people think.

What is critical for evaluation to have an impact, in either case, is that findings be widely disseminated. Vedung (1998) proposes to link instrumental evaluation use to theories of mass media influence on society. He argues that mass media affects society in three domains:

the knowledge domain, the attitude domain, and the action domain. While the knowledge domain concerns perception of reality and the attitude domain pertains to values and valuations, the action domain involves practical action. However, the order in which the domains occur in the real world is subject to dispute. Some scholars argue that the cognitive component (knowledge domain) precedes the affective (attitude domain), which in turn antedates the conative one (action domain). Others believe that action comes first, while attitudes and knowledge change later to legitimize the actions.⁸ (Vedung, 1997: 270)

These communication/persuasion models explain how information can produce knowledge by raising awareness or affecting behavior in specific decisions or action. The role of information, therefore, highlights the key value of dissemination and calls attention to the importance of designing effective and suitable

strategies for the dissemination of evaluation findings. This is of crucial relevance in enhancing evaluation utilization.

The strategy to disseminate the anti-corruption evaluation findings reflects evaluators' concerns about reaching different potential users. Three main axes of dissemination can be identified. First, the final report was extensively distributed within the World Bank, especially during two large communication fairs – the Annual Meetings, October 1998 and the World Bank Knowledge Management Workshop, December 1998. In addition to these 'big-bang' events, evaluators briefed WBI staff on the results of the evaluation whilst it was still in progress. The final report was analyzed in March 1999 at the *ex-post* review session on the anti-corruption program, the meeting where top WBI management assess ongoing activities.

Second, the report was sent to the recipients in Africa, who – as task managers report⁹ – discussed the findings. The report was also distributed to partner institutions internationally and in the field, including Transparency International, the International Federation of Journalists, the Commonwealth Broadcasting Association, the Commonwealth Press Union, Radio Netherlands, the Organization of American States (Trust for the Americas) and networks established by former participants (e.g. the Network of African Parliamentarians Against Corruption).

Third, the report was disseminated amongst the evaluation and social science communities, as for example, during the 1998 Annual Conference of the American Evaluation Association. It was also posted on the WBI website with an executive summary in English and French.

The Evaluation Process: Perceptions and Comments

As stressed in the evaluation report, the study focused on:

- strengths and weaknesses in the development and implementation of anti-corruption activities;
- the impact of such activities;
- opportunities to improve the program.

The evaluation was neither a full-scale impact evaluation, nor a cost-benefit analysis. Rather, the study focused on the reconstruction and assessment of the logic underlying the program. It built on the case studies of the two pilot countries – Uganda and Tanzania – with data gathered through:

- field work (using semi-structured interviews);
- document analysis; and
- a literature review.

The information collected was mainly qualitative; there was no statistical sampling. The amount of data acquired, however, was indispensable in reconstructing how the program had developed. More importantly, the analysis and elaboration of the data enabled the evaluators to propose a set of key recommendations to improve the program in the future.

The key research questions were as follows:

1. What activities have been completed in Tanzania and Uganda during the first three years?
2. On what assumptions (or logic) about national integrity systems (NIS) and limiting corruption is the program based?
3. How were the WBI's activities implemented?
4. What were the impacts of these activities? What information exists about the costs of these activities?
5. What recommendations can be formulated concerning the WBI's anti-corruption program as it expands and increases in scale?

As is apparent, the evaluative questions were designed to illuminate program formulation, process and outcomes. The study had both summative and formative objectives. The study was summative to the extent that it addressed questions of accountability, impacts and outcomes. Of special concern was what the program did or did not accomplish – whether objectives were met, and whether the implementation strategies were successful in moving the program in the desired direction. The study also played a formative role when looking at what kinds of mid-course corrections needed to be made to keep the program on track. The retrospective data gathered during the various stages of the program lifecycle met the task managers' continuing need for information. As the task managers report,¹⁰ their collaborative interactions with the evaluation team helped clarify aspects of the program design and implementation. Although the evaluation study attempted to conceptualize and systematize all information acquired, the task managers state that informal, direct interaction with the evaluation team was the fastest and easiest way of learning the results.¹¹

Whether or not the evaluation was timely depends on the task managers' perceptions. On one hand, task managers recognized the evaluators' sense of timeliness – their attention to what has been called the data shelf life. As they report,¹² the evaluators produced and disseminated information with a conscious regard for their needs. On the other hand, task managers asserted that by the time the report was finalized, the evaluation findings were already outdated because the program had undergone many changes. Such an ambivalent comment casts light on the underlying discrepancy between the practitioners' and users' perceptions of evaluation. There is a disjuncture between the benefits desired by users in the short run and those promised by the evaluators, who are inclined to talk of indirect influences on decision making, of social enlightenment and of cumulative persuasiveness.

The evaluation of the anti-corruption program was timely since it was delivered when the WBI was about to expand the program to include 15 countries. Evaluators have interpreted the informational needs of the WBI as the production of relevant analysis with which to furnish task managers in a useful format and on a continuous basis. Utility, in fact, had been adopted as standard by the evaluation team specifically for the organizational deliberative process. Likewise, as explained below, accumulated learning and enlightenment have occurred over time as a result of the program evaluation.

The evaluation team was partially internal and partially external in order to maximize the advantages associated with the two modes. The internal evaluators had the knowledge of the WBI's culture and decision-making process about what contributes to the success of a program and what hinders better performance. The external evaluators contributed their views on the organization's strategy and added methodological strength to the team (see Kennedy, 1983; House, 1986).

Clearly visible during the research was an attunement of the style of evaluation practice with the character, culture, administrative habits and political realities of the WBI. The evaluators maintained a balance between epistemological considerations and pragmatism.

Example of Instrumental Use: The Media Workshops

One area where the evaluation findings instrumentally fed subsequent program decisions and action was the media workshops. In both Uganda and Tanzania, much of the WBI's effort was concentrated on improving the professional skills of journalists. Towards the end of 1997, six journalism workshops had been organized in Tanzania and seven in Uganda. The workshops were the first in a series of courses that are to last five to seven years and ultimately include nearly all journalists in the two countries.

According to the evaluation findings, interviewees had positive reactions to the materials, personnel and facilities for the courses. They also had a number of criticisms. As mentioned in the final report (Leeuw et al., 1998: 37), participants said that educational materials and case studies should have been more suited to the situation in their countries. They also felt that the materials, especially the videos and the handouts from the World Bank, contained too many references to 'grand corruption'. The 'petty corruption' that is more prevalent in developing countries was virtually overlooked.

Building upon these observations, task managers have begun to put considerable effort into targeting materials to the different regions in Africa. In particular, for all 'Freedonia' exercises, materials were translated into French and Swahili and African cases and simulations are now being collected in association with the Réseau de Partenaires des Médias Africains.¹³ Training materials have also been integrated with local contributions – a local consultant, for example, is in charge of developing a new set of simulation exercises specifically for African countries.¹⁴

The development of more specific training materials is part of a broader effort to further refine the media workshops, whilst the differentiation of training by professional level is meant to reach diverse populations. As the 1999 WBI strategy paper explains, the new planned advanced workshops of six to ten days for more senior journalists will focus on access to information, use of the internet as an investigative tool and interviewing techniques; whereas pilot initiatives of two to three days will be used for editors and publishers (WBI, 1999b).

As underlined in the evaluation report (Leeuw et al., 1998: 38), the first series of workshops was directed solely at print reporters and did not address the specific needs of radio and TV journalists participating. The evaluation findings

highlighted that, especially in the rural areas of both countries, radio is often the most important and effective medium.

As a result, the first generation of media workshops has been replaced by a series of electronic seminars for radio and TV journalists.¹⁵ This mode of delivery ensures far greater coverage, by both geographical area and education level. West United African Television (WUNAT) in particular has taken the lead in broadcasting training modules locally, thus making it possible for the WBI to interact more intensely with organizations based in the regions. This decentralization is of particular importance, as it lays the groundwork for a WBI exit strategy, and for local organizations to take over the training of radio journalists. For the media workshops component of the program, therefore, it is clear that the evaluation findings have helped to shift the program in favor of local organizers and facilitators. Furthermore, the changes that evaluators suggested in the content of materials, format of the seminars and mode of delivery to different audiences helped to tailor the WBI's anti-corruption activities to the local context and to systematically build local capacity. Training materials were revised to make them more Africa-related. The seminar format was modified to encourage participants to be more actively involved and share their experiences. Journalists now participate in the morning sessions and write articles on corruption in the afternoon.

Furthermore, in Uganda, a group of six local organizers and facilitators was trained to continue the media-training program. And in Tanzania, the Media Development Trust Fund was established as an independent organization, working closely with the Association of Journalists and Media Workers (Leeuw et al., 1998). At the same time, the radio seminars were advertised on the radio, in keeping with an evaluator's recommendation to systematically disseminate events to a larger audience and thus increase the impact of the events. In this case, the dissemination of WBI initiatives has been built into the activities themselves.

Another example of instrumental use is that the analysis pinpointed several areas for improvement in program management. According to evaluation recommendations, the transparency of WBI activities had to be increased by using performance indicators. Information about costs was very limited, making it impossible for evaluators to assess the relationship between the costs of the program and its impacts. As task managers report,¹⁶ this observation made them aware not only of the need for more careful budgeting, but also for closely monitoring costs. The financial aspects and cost effectiveness of training are now carefully considered in the decision-making process.¹⁷

Thus, it is clear that evaluation has an impact on decision making at the implementation and first outcome stages of the anti-corruption program. The retrospective analysis of the media workshops has influenced budgeting, the targeting of various populations, the similarities and differences of activities at different sites and aspects of the program that were not operational or not functioning effectively. The mid-term evaluation also helped to identify critical areas of concern and contributed to mid-course corrections in the organization of the media workshops and the mode of delivery. This highlights how evaluation can be a tool for 'program re-engineering'.

Example of Enlightenment Use: The Program Redesign

The reconstruction of the logic underlying the WBI's different activities was one of the major components of the mid-term evaluation. Evaluators applied a theory-based approach that systematically informed the process of data collection and analysis. By eliciting program designers' own theories about how the program was expected to work, they 'disaggregated the assumptions into the mini-steps that are implied and confronted the leaps of faith and questionable reasoning that are (often) involved' (Weiss, 1997: 51).

In general, program goals change in response to changes in political, social, economic and organizational climate, policies, program staff, program structure and clients. Program designers and decision makers operate on the basis of implicit and explicit assumptions, principles and propositions that explain and guide their actions. As a consequence, reconstructing the 'program theory' – a 'specification of what must be done to achieve the desired goals, what other important impacts may also be anticipated and how these goals and impacts would be generated' (Chen, 1990) – may facilitate the understanding of program outcomes for both formative and summative purposes; that is, aid in the decision-making process and highlight the intended and unintended outcomes.

In particular, building on Chen's notion of theory-driven evaluation, the analysis of anti-corruption program theory indicates that it has both prescriptive and descriptive concerns (Chen, 1990). Prescriptive theory deals with what the structure of the program should be in countries requesting the WBI's assistance, and with ways to implement and institutionalize the program. It contains the specific strategies for solving the problem of corruption. Evaluation of programs based on prescriptive theory assesses whether and how the implementation of anti-corruption initiatives differs from its original blueprints. Program staff frequently experience problems in implementing programs since implementation is a very difficult and complicated process. The evaluation of the environment and the overall strategy, therefore, helps program designers and managers to understand whether failures resulted from program design or from implementation.

Descriptive theory, on the other hand, deals with the underlying causal mechanisms that link inputs, implementation processes and outcomes (Chen, 1990). It specifies how the program works by identifying the conditions under which certain processes arise and their likely consequences. It provides an understanding of the program's potential by highlighting the intervening variables, diagnosing potential problems and uncovering causal processes to understand why, for example, anti-corruption activities did or did not work.

In accordance with these two theories, the reconstruction of the underlying logic of the anti-corruption initiatives was presented in two sets in the final evaluation report (Leeuw et al., 1998: 14). The first set shed light on the assumptions about the social, political and economic context in which the anti-corruption program was planned. It reassembled the various components of the WBI's strategy in dealing with the corruption problem by clarifying the goals and outcomes of the intervention. The second set looked into the assumptions about the causal relationships between the anti-corruption activities and their outcomes. In

addition to this reconstruction of the logic, an assessment of its scientific validity was presented. As Leeuw et al. point out, 'such an assessment is necessary, because no evidence exists that a policymaker's (or practitioner, change agent or EDI official's) assumptions are scientifically grounded. However, by the same token, no *a priori* assumption can be made to the contrary' (Leeuw et al., 1998: 14). Thus, special attention was paid to confronting elements of the program's underlying logic with evidence from the social and economic sciences.

The reconstruction of the program logic shows that the WBI program is a strong program, predicated on a set of propositions that can claim a measure of scientific validity. Although a focus is on awareness raising and on changing the cognition and mind-sets of officials, MPs, and others, several mechanisms also steer the activities. Our assessment of these mechanisms reveals potential. In particular, we believe they may reinvigorate social capital and civic society, the sharing and learning processes, and the achievement of transparency and accountability. (Leeuw et al., 1998: 70)

Task managers welcomed this conclusion, which validated the conceptual framework they had developed through intuition and 'learning by doing'. It legitimated their strategy, which in the past had encountered a great deal of resistance. In particular, task managers cite the case of the Bank's legal office denying the authorization for an anti-corruption program on the grounds that anti-corruption concerns were not sound.¹⁸ By searching for empirical evidence and extensively reviewing social science literature, therefore, the mid-term evaluation systematized task managers' implicit thinking and tacit knowledge. It validated the interdisciplinary approach they had built intuitively, based on such concepts as raising awareness of civil society, institutional strengthening, accountability and transparency, empowerment and action research.

As a result, this theory-based evaluation has set the conceptual basis for the program designers' 'alternative' thinking on fighting corruption, as opposed to the traditional economic stance. In addition, the very methodology of evaluation, as opposed to economic analysis, adds empirical evidence to the theoretical framework reconstructed by evaluators. As Picciotto points out:

Evaluators give a privileged role to empirical evidence and make selective use of social science techniques to assess policies and programs. In search of relevant policy recommendations, evaluators prize 'rich description' and manipulate masses of 'dirty' data, whereas economists are inclined to parsimony and limit data collection to the minimum needed to validate 'clean' models. (Picciotto, 1999: 8)

This comment emphasizes, once again, the enlightenment use of evaluation to convey a large volume of methodologically sound information and analysis, which in turn spurs debate on economic, political and social issues, enriches the policy-making process, complements analysis from other disciplines and contributes to knowledge building and sharing.

Another aspect of the enlightenment use relates to the influence that a specific evaluation finding can have on task managers' thinking and redesign of the anti-corruption program. As the evaluators state in their report:

There are also weaker points in the logic. One concerns the emphasis on awareness raising . . . People can expect no automatic progression from awareness of an unjust

situation to intervening to bring it to an end. Another is the belief in empowerment. In our review of the research literature, we did not find evidence that this mechanism will indeed be effective. Even when individuals are empowered, it is not certain that empowerment at the social or organizational levels will follow. When a program aims at empowering larger groups of people, different levels of empowerment must be taken into account. Empowerment at one level does not necessarily lead to empowerment at another. Finally, it was pointed out that the prospects of the workshops' contents trickling down to society at large are not particularly good. UNDP data show that the necessary communication infrastructure is not very well developed in Uganda and Tanzania. (Leeuw et al., 1998: 70)

These remarks seem to have had significant bearing on the recent redesign of the anti-corruption activities. The new WBI strategy paper, 'Controlling Corruption: Toward an Integrated Strategy,' states that:

Traditional anti-corruption courses were based on two different approaches to the problem of corruption. One type emphasized the analytical understanding of the problem; that is, corruption is ultimately a symptom of weak institutions and poor policies. Thus, addressing corruption effectively means addressing underlying economic, political, and institutional causes. The second type of course was more proactive, dealing with the process of awareness raising, mobilization and civil society involvement in the fight against corruption.

The new so-called 'core courses' combine the two approaches into an integrated framework. This key component of the courses is to provide the participants with the necessary tool kit to enable them to design a coherent anti-corruption strategy that is tailored to their country's specific institutional and political realities. Through a series of interactive sessions, participants will work through the process of designing an anti-corruption strategy and discuss the challenges of integrating the participatory process with concrete institutional reforms. (WBI, 1999c: 3)

Conclusions

The present study has shown, through a series of examples, how evaluation has been used in both an instrumental and an enlightenment fashion. It can be concluded that links between knowledge generation and utilization are not always clear and direct, and that specific information cannot be isolated as the basis for a particular decision. Nevertheless, the examples show that utilization has occurred and that this in turn brings about change in program design and implementation.

At the same time, the instrumental and enlightenment perspectives were applied to identify other potential users, although the article focuses only on task managers' utilization. In this regard, the role of dissemination of evaluation findings to enhance utilization has been explicitly reconstructed. As the analysis has shown, evaluators have sought to communicate and create awareness of evaluation findings through conferences, workshops, the internet, professional media, institutional meetings and policy networks. The best ways to encourage the use of evaluation findings have been to involve the program staff in defining the study and helping to interpret results, and to produce regular reports for the

program staff whilst the study is in progress. As Weiss (1998b) comments: 'this kind of sustained interactivity transforms one-way reporting into mutual learning' (p.30).

As evaluation in this domain involves the production of knowledge concerning the effectiveness and efficiency of development interventions, a participatory style enables people to share valuable information and analytical capacity. More importantly, close and collaborative relations among program designers, implementers and evaluators ensure that evaluation results are utilized to improve daily practice and also to make larger changes in policy and programming.

Acknowledgements

I would like to thank Ray Rist who supervised the original report, and also Frans Leeuw, principal evaluator of WBI Anti-Corruption Initiatives, who provided me with helpful sources of reading. In addition, I am grateful to Rick Stapenhurst for the interviews he granted me and for his prompt feedback and valuable suggestions. Finally, I am especially grateful to Olivier Butzbach who patiently read the manuscript and offered precious advice.

Notes

1. This article draws on a World Bank Institute (WBI) working paper on the utilization of evaluation (WBI, 1999a) which the author completed while she was Evaluation Analyst within the Evaluation Unit of the World Bank Institute. However, the opinions expressed here are totally personal. They reflect neither the view of WBI task managers nor the position of the World Bank.
2. Formerly Economic Development Institute (EDI).
3. A revised version of this report was published by F. Leeuw, G. H. C. van Gils and C. Kreft with the title: 'Evaluating Anti-Corruption Initiatives: Underlying Logic and Mid-Term Impact of a World Bank Program' (Leeuw et al., 1999). Yet the present article refers exclusively to the original WBI evaluation, fully cited in the Reference list (Leeuw et al., 1998).
4. For another WBI Evaluation Utilization Study see Marra, 1999.
5. The Freedonia simulation is a continuously unfolding story in the fictitious land of Freedonia, with four newspapers competing to get the best stories. Each day the Newsroom receives information on possible corrupt practices, and journalists have to decide what the reports mean, what lies behind the story, what line or approach to take and how to write stories accordingly.
6. As Argyris and Schon pointed out some years ago, those lessons will have little impact on collective behavior or on the effectiveness of programs unless they become part of the 'organization culture' and are incorporated into the models and program theories that govern how resources are used and field activities carried out.
7. 'Whenever an error is detected and corrected without questioning or altering the underlying values of the system (be it individual, group, inter-group, organizational or inter-organizational), the learning is single-loop. Single-loop learning occurs when matches are created, or when mismatches are corrected by changing actions. Double-loop learning occurs when mismatches are corrected by first examining and altering the governing variables and then the actions . . . Single-loop learning is appropriate

for the routine, repetitive issues – it helps get the everyday job done. Double-loop learning is more relevant for the complex, non-programmable issues – it assures that there will be another day in the future of the organization’ (Argyris, 1994).

8. This is the model of dissonance-attribution theory. According to this perspective, learning occurs in the reverse order: first comes behavior, then attitude modification, and finally learning. While admitting that information activities can have a cognitive effect in promoting the original choice behavior and attitude change, dissonance theorists contend that the main information effect is in terms of reducing dissonance or providing information for attribution or self-perception after action and attitude transformation have occurred (see Vedung, 1998).
9. Based on interviews with WBI task managers.
10. Based on interviews with WBI task managers.
11. Based on interviews with WBI task managers.
12. Based on interviews with WBI task managers.
13. Based on interviews with WBI task managers.
14. Based on interviews with WBI task managers.
15. Based on interviews with WBI task managers.
16. Based on interviews with WBI task managers.
17. Based on interviews with WBI task managers.
18. Based on interviews with WBI task managers.

References

- Argyris, C. (1994) *On Organizational Learning*. Malden, MA: Blackwell.
- Argyris, C. and D. A. Schon (1978) *Organizational Learning*. Reading, MA: Addison Wesley Publishing Company.
- Chen, H. (1990) *Theory-driven Evaluation*. Newbury Park, CA: Sage Publications.
- Hirschman, A. O. (1967) *Development Projects Observed*. Washington, DC: Brookings Institution.
- House, E. (1986) ‘Internal Evaluation’, *Evaluation Practice* 7(1): 12–47.
- Kennedy, M. M. (1983) ‘The Role of the In-House Evaluator’, *Evaluation Review* 7: 519–41.
- Leeuw, F., G. H. C. Van Gils and C. Kreft (1998) *EDI's Anti-Corruption Initiatives in Uganda and Tanzania – A Mid-term Evaluation*. Washington, DC: World Bank Institute.
- Leeuw, F., G. H. C. Van Gils and C. Kreft (1999) ‘Evaluating Anti-Corruption Initiatives: Underlying Logic and Mid-Term Impact of a World Bank Program’, *Evaluation* 5(2): 194–219.
- Marra, M. (1999) ‘How Much Does Evaluation Matter? A Follow-up on the Tracer Study on Training-of-Trainers Seminars in Africa’, *WBI Evaluation Studies* 1(June).
- Pawson, R. and N. Tilley (1997) *Realistic Evaluation*. London: SAGE Publications.
- Picciotto, R. (1999) ‘Towards an Economics of Evaluation’, *Evaluation* 5(1): 7–22.
- Rist, R. C. (1999) *Program Evaluation and the Management of Government*. New Brunswick, NJ and London: Transaction Publishers.
- Rossmann, G. B. and Rallis, S. F. (1998) *Learning in the Field*. London: SAGE Publications.
- Vedung, E. (1997) *Public Policy and Program Evaluation*. New Brunswick, NJ and London: Transaction Publishers.
- WBI (1999a) ‘WBI's Anti-Corruption Activities: Evaluation's Impact on Program Implementation and Redesign’, working paper, *WBI Evaluation Studies* (October 1999).
- WBI (1999b) *Activity Brief for Fiscal Year 1999–2000*. Washington, DC: World Bank.

- WBI (1999c) 'Controlling Corruption: Toward an Integrated Strategy', strategy paper, Washington, DC: World Bank.
- Weiss, C. (1997) 'Theory-based Evaluation: Past, Present, and Future', *Progress and Future Directions in Evaluation: Perspectives on Theory, Practice, and Methods* 76 (Winter): 41–55.
- Weiss, C. (1998a) *Evaluation*, 2nd edn. Upper Saddle River, NJ: Prentice Hall.
- Weiss, C. (1998b) 'Have We Learned Anything New About the Use of Evaluation?', *American Journal of Evaluation* 19(1): 21–33.

MITA MARRA is a doctoral student at the George Washington University. She has worked in the Evaluation Unit of the World Bank Institute since January 1998. She is currently consultant for the Operations Evaluation Department (OED) of the World Bank working on the OED Anti-Corruption and Governance Study. She has worked within the Italian government on the pilot project '100 Initiatives at the Service of Citizens' for the reform of public sector. She is interested in the role of evaluation in public sector reform. Please address correspondence to: The World Bank, 1818 H Street, NW, Washington, DC, 20433, USA. [email: Mmarra@worldbank.org]