

CHAPTER-BY-CHAPTER MANAGEMENT OBSERVATIONS ON METHODOLOGICAL AND INTERPRETATION ISSUES

Chapter 1

Figure 1.3 is based on WBI training programs in 13 focus countries, FY02–06. Paragraph 7, however, presents the information as though it represented all of WBI training.

Chapter 2

Client Survey Response Rate

Much of the evaluation's discussion of outcomes is based on a survey of former participants in events that were sponsored or cosponsored by either WBI or Bank-funded projects in six countries. An earlier version of the evaluation reported that IEG's contractor had to contact approximately 2,500 participants to obtain 548 responses, yielding a response rate of about 22 percent. Of these respondents, IEG excluded 37 because they had not stayed in their job for at least six months after the training.

Despite the low response rate to the survey, the evaluation provides findings disaggregated by country (for example, paragraph 9 on Mexico), without attention to margins of error or the representativeness of the sample.

IEG reports an average response rate of 33 percent (averaging across the six countries) in the final report, but the country-level response rate varies from 11 percent in Mexico to 64 percent in Burkina Faso. In three other countries the response rate varied from 21 to 30 percent. In Tunisia, a response rate of 47 percent was achieved but with only 39 respondents. The raw response rate in the six countries together was about 23 percent, and 7 percent of those respondents were dropped because they had not stayed in their jobs for six months after training and reported that the training was not relevant for their current job. With this response rate, the danger of a bias calls for caution in interpreting the results. To avoid these biases many accepted professional norms call for higher response rates. The minimum acceptable response rate for surveys conducted for the U.S. National Center for Education Statistics is 85 percent. The U.S. Federal Government's *Standards and Guidelines for Statistical Surveys* indicate that surveys should be designed to achieve a "target response rate of at least 95 percent, or provide a justification for a lower anticipated rate." When IEG sponsored an evaluation of its client training program, a response rate of 39 percent was declared to "impose constraints on the conclusions" and did not permit generalizations across all participants.¹

Management also has concerns about the wording of several questions in the surveys of clients and task team leaders. For example, the report fails to note that 14 percent of

¹ Evaluation of the International Program for Development Evaluation Training. Jua Management Consulting Services, March 22, 2004.

WBI's respondents participated in online events and another 13 percent in videoconferences. As a consequence, a portion of these participants may not have had opportunities to interact with other participants. Despite this situation, the IEG survey asked these participants whether they learned from the experience of other participants in the course; what their was judgment about the quality of lectures, when there may not have been any lectures; and whether participants had equal levels of knowledge coming into the course.²

Proportion of Client Survey Respondents Noting Positive or Greater Change in Their Work

Paragraph 9 states that country statistics on workplace outcomes ranged from a low of 37 percent of WBI respondents in Mexico to a high of 90 percent of project respondents in Nigeria. However, when respondents were asked to reconfirm their responses (in question 6d of the survey) 45 percent to the respondents in Mexico noted that the course had "resulted in substantial positive changes." Responses to the same survey question (6a) also show that 98 percent of WBI respondents in Mexico reported that their participation in a WBI event had led to positive or greater change in their work.

The text in this chapter also uses the terms "workplace performance changes," "outcomes," and "impact on key work functions" interchangeably, whereas the last term implies a substantially different result from the first two.

Chapter 3

IEG's evidence does not support the claim that "WBI's training performs much more poorly on targeting of training" (see Evaluation Highlights at beginning of chapter 3). In fact, the data in table 3.3 reveal no difference between project-supported and WBI training. Table 3.5 further illustrates that WBI and project-supported training were equally good (or deficient) in selecting participants who had about equal level of knowledge coming into the course, a factor that was found to be predictive of training outcomes.

Table 3.2 states that approximately half of the respondents responded that the course covered the "...right amount of topics for the amount of time allotted." The survey question in fact asked respondents to agree or disagree with the statement, "The course covered too many topics for the amount of time allotted." Thus, disagreement with this statement could also indicate that the course length was too long and the topics too few for the time allotted.

Paragraph 10 states that the average length of training in the field reviews was less than five days, regardless of the course objectives. However, the evaluation provides no

² IEG notes that both online and videoconference courses provide opportunities for dialogue and discussion and a large percentage of WBI online courses also have face-to-face components. Moreover, survey respondents also had the option of answering "not applicable" to this question.

information on whether the training was linked to other training activities and capacity-building efforts. The evaluation then compares the average duration of Bank-supported activities with that of other external training, regardless of the objectives of the training. Furthermore, how the length of the training is calculated depends on how the activities are coded and accounted for by different organizations.

Paragraph 16 provides benchmark information on follow-up support to trainees in Motorola University's learning program, but does not note that this is a learning program for the company's employees. Training of staff for a company that manufactures products is far different than assistance provided to enhance capacity in developing countries.

Table 3.3: that about 75 percent of respondents indicated that the training addressed their country's circumstances and addressed issues important to their work seems quite respectable. Indeed, this high percentage seems inconsistent with the finding in paragraph 23 that "only two out of eight of WBI program reviewed had adequate diagnosis."

Paragraph 30 reports field studies suggesting that participant selection was better in project-supported training than in WBI training. However, the data in table 3.5 show that that a majority of respondents learned from the experience of other participants in the course.

Chapter 5

Use of WBIEG Findings

In discussing WBI's collaboration with country teams, chapter 5 highlights and uses WBIEG's findings that are critical of WBI, but does not include favorable findings in its analysis. For example, WBIEG's survey of all focus country teams revealed that 98 percent of respondents said WBI satisfactorily consulted with the country teams to select its activities. IEG notes this finding but then states that "WBI's overall work program has not led to systematically better collaboration on the level of task managers and operational task teams," without any acknowledgement of the fact that the program may be designed in response to specific requests from the country team and not necessarily for specific task managers (paragraphs 22 and 23). In contrast, the IEG report notes WBIEG's finding that WBI's programs are insufficiently coherent (paragraph 25).

Analysis of, and Recommendations with Respect to, M&E Requirements

In its analysis of the M&E systems of the training programs reviewed, the IEG report seems to suggest that performance outcomes and indicators of performance outcomes are desirable for all training activities, and that all such activities are amenable to all levels of the Kirkpatrick model of evaluation (e.g., paragraphs 6 and 9). Management disagrees and notes that certain types of training activities are not amenable to this framework.

Management notes that WBI *monitors* the number of participant training days and participant satisfaction at the program and activity levels. These data provide early, but not definitive, indications of training weaknesses. WBI *evaluates* the outcomes of WBI training on the basis of representative samples of activities and participants. These evaluations provide information to WBI management for accountability purposes and to WBI task managers for learning with respect to effective features of training design and implementation. However, since the evaluations are based on samples, they are insufficient for diagnosing performance issues with respect to, and outcomes of, specific activities.

With regard to box 5.1 and paragraphs 11 and 12, Management does not agree that the focus of WBI level-1 evaluation is limited and captures only participant satisfaction. The instrument asks about the utility, relevance, and usefulness of the training. It is appropriate for the goals of level-1 evaluation, which are to obtain immediate feedback from participants inexpensively and quickly for improving future deliveries of the capacity-building activity. A recent external review of WBI's assessment practices, which is cited in the review, concluded that WBI's level-1 evaluation surveys "are among the most complex and sophisticated questionnaires currently in use."³ This conclusion is based on the reviewer's consideration of assessment practices in bilateral and multilateral development agencies, universities offering professional development programs, and several private sector firms conducting staff programs, including Motorola University (one of the IEG report's benchmark organizations), Caterpillar University, and Knowledge Advisors. This review further noted that the Inter-American Development Bank and the U.S. Agency for International Development currently conduct only level-1 evaluations because "small budgets have limited their ability to conduct higher level evaluations of training."

Management notes that the report seems to suggest that assessments of participants' reaction to training (commonly referred to as level-1 evaluation) is useless. Yet IEG espouses the Kirkpatrick model, which includes level-1 evaluation as a key component. WBI uses level-1 data as a way of obtaining quick participant feedback. Furthermore, level-1 data are obtained for specific courses and sessions, and the information is provided to the task manager of that particular course or session.

The report notes that its survey of benchmarked international training organizations [available on the report's Web site at www.worldbank.org/ieg] was used to "provide examples of good training management practice" (paragraph 24). Although the report questions the value of level-1 evaluations, five of the six benchmarked organizations use level-1 evaluations for their courses. It is unclear whether the sixth organization uses level-1 evaluation. According to IEG's report, the International Training Center of the

³ Evans, David. 2007. *Are There New Approaches to the Evaluation of Learning Programs?* Report No. EG07-125 World Bank Institute, Washington, DC. Last year the Bank's Office of the Chief Learning Officer (CLO) asked the Corporate Executive Board's Learning and Development Roundtable to review the report, and to provide an assessment of the report's quality. Here is what the Board told the CLO: "Overall, the report is an excellent summary of the state of learning evaluation. It provides a thorough listing of the most commonly used evaluation approaches and, based on the information provided in the report, offers an accurate assessment of WBIEG's [approach to] evaluation as one of the best."

International Labor Organization is an example of a “best-practice” organization, but it does not conduct any level-2 evaluations and does not have the resources to conduct level-3 evaluations. Within MASHAV, according to the report, level-3 evaluations are sporadic. Finally, it is worthwhile to note as well that IEG itself promotes the use of level-1 evaluations in the training that it sponsors (for example, in its International Program for Development Evaluation Training).

Paragraph 14 notes that “WBIEG has mostly discontinued level-3 evaluations of specific sector and thematic programs.” This is an inaccurate statement, as noted earlier. The report’s assertion that “responses provide information on a country basis but not on a program basis” is also inaccurate, given the program-level evaluations reported in the review itself.

Figures 5.2 and E.1 appear to be inconsistent with the number of TTLs (43) interviewed.

Paragraph 23, last bullet point, states that 10 of the 20 TTLs surveyed reported that “...the relationship had stayed the same, *with no improvements*” (italics added). In fact the survey response category was simply that the relationship had “stayed the same.” Thus adding the words “with no improvements” implies that the relationship needed improvements, when the relationship might already have been strong and productive.