THE INFLUENCE OF TRAINING EVALUATORS' CHARACTERISTICS ON EVALUATION PRACTICES IN MALAYSIA

Junaidah Hashim

Assistant Professor, Department of Business Administration, Kulliyyah of Economics and Management Sciences, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur (email: junaidahh@iiu.edu.my)

ABSTRACT

This paper examines the background characteristics of training evaluators and the way training evaluation is conducted in Malaysia. The background characteristics examined include evaluators' current involvement in evaluation activities, training and experience, academic qualification, and familiarity with evaluation models. This study adopts an instrument developed by Shadish and Epstein (1987). Ninety-four training evaluators participated in this study. The findings reveal that training evaluators are active in evaluation activities, receive informal training in conducting evaluation, possess academic qualifications in various disciplines, experience limited training evaluation and have a low level of familiarity with evaluation models. The evaluators' background characteristics are found to have influenced the way evaluation is conducted. The implications of the findings are discussed.

JEL classification: I2, M1
Key words: Training, Evaluation practices, Malaysia

1. INTRODUCTION

Training programs are now an essential feature of organizational life. Organizations have come to realize the importance of training as a factor for organizational growth and survival. Many companies in the United States are known to have spent as much as USD40 billion a year on training programs for managers and executives (Mann and Robertson, 1996).

Such investment in training is not limited to the USA. Training importance applies equally to Malaysia. The Malaysian government, since independence in 1957, has manifested its commitment toward human resource development. The Malaysian government has not only increased the budget for training in each of five-year plan (RM400 million for Eighth Malaysia Plan as compared to RM223.7 million for the Seventh Malaysian Plan) but also has established a special agency to monitor the training activities in the country. At the same time, the government has introduced the Human Resource Development Act 1992, which requires a company to contribute a one-percent equivalent of its monthly payroll to the Human Resource Development Fund which would then be used to promote training for the employees. Organizations have always been concerned about the effectiveness of their training programs. They are concerned about the value for money they get from their training budgets. One of the best ways organization can find out about either of these concerns is through a systematic process of evaluation.

1.1 PROBLEM STATEMENT

Researchers and trainers alike generally agree that training evaluation is an important part of the training system (Geertshuis et al., 2002). However, the practice of evaluation in training receives much criticism in the field's literature. It was found that training evaluators tend to use different methodologies and approaches from the prescription given in the literature. The issues in training evaluation center around which training outcomes are important and how they should be measured. Training evaluators are criticised for having used unsystematic and simple methods in evaluating the training programs. Many training evaluators are said to have lacked the ability, skill, knowledge, and unbiased viewpoint about evaluation that are necessary to perform one (Shamsuddin, 1995; Chen and Rossi, 1992; and Shadish and Epstein, 1987).

1.2 RESEARCH OBJECTIVES

The objective of the study is to identify who the training evaluators in Malaysia are. Specifically the study addresses the following questions:

1. What are the background characteristics of training evaluators in Malaysia?
2. How do the training evaluators conduct the evaluations for the program they delivered?
3. Do the background characteristics of training evaluators influence the way the evaluation is conducted?

1.3 SIGNIFICANCE OF THE STUDY

From a theoretical perspective, the findings would provide a valuable knowledge base for evaluation theory. The empirical information of this study would provide greater insights as to how and to what extent the involvement, training, knowledge and the understanding of evaluation models actually influence the actions of training evaluators. It would inform us about the complexities faced by training evaluators in trying to apply evaluation models in a real practice situation.

2. REVIEW OF LITERATURE

Evaluation is a systematic process of collecting and analyzing information for and about a program which can be used for planning and as a guide to decision-making as well as assessing the relevance, effectiveness and the impact of various program components. Evaluation in training is an elusive concept especially when it comes to practice. There still appears to be more talk than action. Training providers seldom attempt to determine the effect of their program.

2.1 THE PRACTICE OF TRAINING EVALUATION

The basic intent of any evaluation is to get direct feedback (McClelland, 1994). Evaluation can serve as a useful component of all programs and services, ideally in a formative sense to make amendments and improvements as well as a general monitoring role (Athanasou, 1998). In a more strategic sense, evaluations can also be devised to measure long-term effects such as what learning and behavioral changes have occurred (McClelland, 1994).

   According to Talbot (1992), many organizations have either ignored

or approached evaluation of training in an unconvincing or an unprofessional manner. This concurs with the findings of Maimunah (1990) and Shamsuddin (1995) which showed that the evaluations conducted by the extension agencies are incomprehensive, *ad hoc*, informal and unsystematic. Comprehensive evaluation refers to evaluating a program as a whole through inter-related evaluation activities and the proper use of evaluation results, while systematic evaluation refers to a well-planned and a well-conducted evaluation approach, and the utilization of proper research procedures to collect data and analyze the data.

Chen and Rossi (1992) also commented that evaluation knowledge found in the literature is not being fully utilized in training evaluation practices. There are more than fifty evaluation models available; however the evaluators tend to use the four-level Kirkpatrick Model (Phillip, 1991; Geerthuis et al. 2002).

Kirkpatrick (1994) proposes four levels of evaluation: Reaction (assessing what the participants thought of a particular program), Learning (measuring the learning of principles, facts, skills and attitudes), Behavior (measuring changes in aspects of job performance), and Results (changes in criteria of organizational effectiveness).

The most common way of evaluating training, whenever it is done, is through the reaction of the participants to the training (Rosti and Shipper, 1998). Evaluation at this level is associated with the terms 'smile sheet' or 'happiness sheet.' Participants' reactions are easy to collect, but provide little substantive information about training effectiveness and worth (Shelton and Alliger, 1993).

2.2 BARRIERS TO EFFECTIVE TRAINING EVALUATION

According to Mann and Robertson (1996) many researchers believe that one of the main barriers to employing effective evaluation procedures for training programs is the difficulty of knowing what and how to evaluate. The question of what to evaluate is crucial to the evaluation strategy. As pointed by Redshaw (2000) factors other than training such as marketing activities, market forces, and new technology may have some influences on the organizational effectiveness. Another difficulty is that the results of training may take considerable time to show up in overall results.

The evaluators tend to evaluate the training at the lowest level of the model (reaction level). This is a matter of convenience as evaluation during this level is easier to carry out than a follow-up evaluation (Bramley, 1999). Training evaluators are reluctant to conduct Level 4 evaluation (result level) of the Kirkpatrick Model because it is time consuming (Dixon, 1996) and the process is complex. Training evaluators prefer to use a simple method because it is easy to perform one and its process takes a shorter time to complete. Typically, most employees' training is only assessed by the individuals experiencing the process, and making use of questionnaires and verbal responses. This technique can be considered to be highly subjective (Adams & Waddle, 2002). Many authors agree that trainers often do not have the skills and ability to conduct evaluations as they may be direct line managers and have never been trained formally (Garavaglia, 1993; and Shamsuddin, 1995).

3. METHODOLOGY

3.1 SUBJECTS

This study involves the training institutions which have registered with the Human Resource Development Council as training providers for Approved Training Program (ATP/PROLUS) scheme. The training providers of the Council are an appropriate sample to represent the training evaluators in Malaysia. It is a requirement for all training institutions to register as official training providers with the Council to enable their clients to get reimbursement for their training expenditures. The researcher has obtained a list of training providers from the Council and based on the list there are 279 registered training providers. However, this list was prepared in 1998 and it has not been updated since then by the Council. When this study was conducted in 2000, the researcher has found that 17 training providers were no longer operating at the registered addresses. Thus, these providers were eliminated from the total sample. The actual sample size was only 262 training providers.

2.2 INSTRUMENTATION

This study uses a mailed questionnaire due to the scattered distribution

of the sample all over the country. The questionnaire was adopted from a previous research done by Shadish and Epstein (1987) with some minor adjustments to suit the local context. The adjustment made involves omission of three variables from the original questionnaire. The omitted variables are: factors causing changes in evaluation practice, perceived resource constraints in evaluation, and why concepts are influential. One new variable (Evaluation Schedule) was introduced in the questionnaire for the study.

The questionnaire was designed to be completed by an individual responsible for evaluating training in each training institution. The questionnaire consists of two parts. Part A seeks background information about the evaluators. It gathers data about the training evaluator's nature of current involvement in evaluation, his highest academic qualification, his field of academic training, his type of employment setting, his training experience in evaluation, his opinion on the importance of various training experiences on his evaluation practice, and his familiarity with evaluation models. Part B collects data about the evaluation practices. It inquires information about the evaluation methods used, activities done to facilitate the use of evaluation findings and the evaluation schedule.

3.3 PRE-TESTING OF THE RESEARCH INSTRUMENT

Since there are some modifications to the original questionnaire, a pre-test was conducted to assess the appropriateness and practicality of the instrument. Questionnaires were distributed to 20 individuals whom the researcher knew to have some experience in both designing and evaluating training. The Cronbach's coefficient alpha values for the pre-test and the final stage of selected variables are displayed in Table 1. There were some items that had to be deleted in order to increase their standardized Cronbach alphas. For the 'evaluation practice' variable, the initial standardized item alpha was quite low; it was 0.3655. Two items (EP3: to judge program value, and EP4: to explain how the program worked) were deleted. After deletion, its standardized alpha increased to 0.7002. In the final questionnaire, EP3 was deleted, whereas EP4 was rephrased.

The validity of the instrument used in this study was done by face validity. Face validity means either the variables or the instrument has

been evaluated by an expert. Based on face validity, the instrument used in this study is acceptable because it has been used in an earlier study done by Shadish and Epstein (1987).

3.4 DATA COLLECTION

A telephone call was made to an initial contact person at each institution. The study was described to this initial contact person, and if necessary, a more appropriate representative of the institution was identified to be a respondent. The questionnaire was then mailed to the identified person to be completed. Each questionnaire was coded for the purpose of keeping track of non-responses, so that additional requests could be made if necessary.

A reminder letter was sent to the identified persons when the questionnaire was not returned within 14 days after the initial mailing. The non-respondents were given another seven days to mail the completed questionnaires. On the seventh day, a follow-up telephone call was made to remind them again. When necessary, another copy of the questionnaire was sent to improve the response rate. Respondents were given the option to fax the questionnaires to speed up the data collection process.

When the questionnaire was first distributed, 64 responses (24 percent) were received. After a reminder was sent, another 13 responses were received (increased to 29 percent). When telephone calls were made, another seven responses were received (increased to 32 percent). To further improve the response rate, the researcher personally collected some of the questionnaires from the non-respondents who were located in Kuala Lumpur and managed to collect another 13 responses. The researcher then started the data analysis when there was no mail at all received for the next seven consecutive working days. The seven-day period with zero mail received was a good indicator that there would be no more responses to be received. Overall, a total of 97 subjects (37 percent) responded. However, three responses were discarded because the respondents did not fulfill the requirements as respondents, as they were not involved in any of the evaluation activities listed in the questionnaire. As a result, there were only 94 valid subjects (36 percent) used in the data analysis.

The researcher called 35 selected training institutions that did not

respond. The purpose was to ask questions for a few vital variables in the questionnaire. These 35 telephone responses were treated as a separate subject. Specifically, it was meant to examine whether there was any significant difference in the mean of the selected variables between the respondents and non-respondents' responses. The data obtained through the telephone interview were analyzed by using an independent *t*-test. The result shows that there was no difference in mean of selected variables (evaluation purpose *?*=0.628, evaluation schedule *?*=0.417, and evaluation method *?*=0.283) between these two groups.

4. FINDINGS AND DISCUSSION

4.1 BACKGROUND CHARACTERISTICS OF TRAINING EVALUATORS

a. Current involvement in evaluation

evaluators, teaching evaluation, and academic interest in evaluation. Respondents were asked to answer either "yes" or "no" to each of these activities. The respondents' highest involvement was in using evaluation activities (90.4 percent), and the lowest involvement was in teaching evaluation (33.0 percent).

b. Training experience in evaluation

Table 3 depicts the distribution of respondents' training experience in evaluation. Specifically, it shows how many of them have attended an evaluation course and how many of them have been specifically trained to become an evaluator. A total of 43.6 percent of the respondents indicated that they have attended an evaluation course, while 11.7 percent indicated that they have been trained as an evaluator.

c. Current employment setting

Table 4 displays the employment setting in which the respondents are currently involved. Training providers participating in this study included government agencies, private training institutions, private

firms, colleges and universities, and others. However, private training institutions were the biggest training providers in Malaysia (54.3 percent) participating in this study.

d. *Academic qualification*

As shown in Table 5, the majority of the respondents were graduates. There were 2.1 percent who hold doctoral degrees in their fields; 27.7 percent have a master's degree, 51.1 percent have a bachelor's degree, 14.9 have diplomas, and 4.3 percent have Higher School Certificates.

e. *Academic field*

A few relevant academic fields were identified to find out in which area respondents had been academically trained. The academic fields identified were evaluation, business administration, psychology, education, and others. There was one respondent whose academic qualification is evaluation. The majority of the respondents come from other academic fields; 26.6 percent come from business administration, 25.5 percent come from education, and 5 percent come from psychology. The "other" 41.5 percent category consists of respondents from either engineering or computer science backgrounds.

f. *Importance of training experience in evaluation*

This part sought information about how important various training experience are to the respondents in conducting their evaluation. Table 7 shows that respondent's own effort was considered to be the most important experience (mean=4.15). The second most important experience was discussing with clients (mean=3.92). The next experience that was considered to be tangentially important to respondents was reading books and articles (mean=3.83). Respondents had also attended occasional academic courses but feel that this experience was totally unimportant to them in conducting an evaluation.

g. *Evaluation model familiarity*

Respondents were asked to indicate their familiarity with 19 evaluation

models commonly found in the field's literature. This is basically aimed to find out which evaluation model respondents are familiar with and to assess their overall familiarity with evaluation models. As shown in Table 8, the most familiar model is Comparative evaluation model by Campbell and Cronbach. There were 58.5 percent of the respondents who recognized or knew this model. The next familiar model is Kirkpatrick's four level model. A total of 41.5 percent of the respondents knew at least something about the model. All the responses in model familiarity were computed as a total score. This was done to determine the overall model familiarity among the respondents. The result showed an overall mean of 1.27, which is regarded as low (do not recognize the model).

4.2 THE PRACTICE OF TRAINING EVALUATION

a. *Evaluation methods used*

Respondents were asked to indicate how frequent each evaluation method listed was used in the evaluation they conducted. The results show that the respondents used all the methods commonly found in the literature. However, as shown in Table 9, trainee feedback was found to be the most frequently used evaluation method (mean=4.19). Other frequently used methods were observation (mean=3.84), and interview (mean=3.65).

b. *Activities to facilitate the use of evaluation findings*

Table 10 displays eight activities used to facilitate the use of the evaluation results. Respondents were asked to indicate their agreement or disagreement with respect to each of them. Three activities were found to be used most by the respondents. Evaluation results were used to provide feedback to improve program planning (mean=4.56), to disseminate a written report of results (mean=4.22), and to translate results into recommendations (mean=4.14). Since respondents conduct training for clients, providing oral briefings to clients was also important (mean= 3.89). However, the respondents did not seem very keen to publicize the evaluation results in the media (mean=2.39).

c.   *Evaluation schedule*

Pertinent literature indicates that systematic evaluation should be conducted at the planning phase, during the implementation phase, after training was conducted, and a few months after the training has been conducted. Table 11 shows the percentage of responses to each of the statements. Almost 90 percent of the respondents agreed that they evaluate their training immediately after the training is completed. Evaluation during the implementation was the second most frequently conducted type of evaluation. Each statement that was relevant to comprehensive evaluation and systematic evaluation was then computed as total scores. The total score for comprehensive evaluation and systematic evaluation were 3.70 and 3.72, respectively. This shows that the evaluators have conducted a moderately comprehensive and systematic evaluation.

3.3 THE INFLUENCE OF EVALUATORS' BACKGROUND CHARACTERISTICS AND MODEL FAMILIARITY ON EVALUATION PRACTICE

To determine the respective influence on evaluation practice, evaluators' background characteristics and model familiarity were treated as separate variables. Correlation analysis shows that there is a significant positive correlation between background characteristics and evaluation practice (r=0.459, ?=0.000), and there is also a significant correlation between model familiarity and evaluation practice (r=0.393, ?=0.000. Based on regression analysis both variables, i.e., the evaluators' background and the evaluators' model familiarity are also found to have a significant positive relationship with evaluation practice. However, based on these results, it was found that the evaluators' background is a better predictor for evaluation practice. The details are shown in Table 12 and 13.

The actual practice of training evaluation does not often follow the methods and procedures as found in the literature. This digression is largely explained by the fact that many training evaluators have not found the literature's advice applicable to their organizations. This probably explains why the evaluators tend to use simple methods such as trainee feedback. They have not formally been trained on how to

conduct an evaluation. They learn about conducting an evaluation from their own past experience, discussing with clients, and reading books and articles on their own. They do not know what the literature has to offer them. So, it is rather difficult to modify the available evaluation models to suit their setting and requirements. Had the evaluators been trained formally, they would probably be able to make effective use of evaluation models.

Although the evaluation methods used by the evaluators do not meet the traditional academic notion of systematic evaluation, it does not necessarily mean that the practice of training evaluation has been conducted in an unconvincing or an unprofessional manner. What practitioners require is an evaluation that is done in such a way as to provide valuable information to meet practical needs, is reproducible, and can be quickly and easily conducted. As revealed in the findings, the evaluators do, to a certain extent, conduct moderately comprehensive and systematic evaluations although the method used is simple. In terms of facilitating use of evaluation results, the evaluators have used the results for providing feedback to improve program planning, disseminating a written report, and translating the results into recommendations. This is consistent with what have been suggested in the literature.

However, as suggested by Caffarella (1994), it is important that the practitioners who are involved in evaluation have not only the working knowledge about evaluation approaches, but also the knowledge about data collection techniques, data analysis and program judgments. This probably will help the evaluators to adopt a more complex evaluation method whenever necessary with the hope of producing better evaluation results. This knowledge does not necessarily have to be obtained through formal academic training; however, practitioners can learn through reading and discussing it with other practitioners. What matters is that practitioners must have some basic knowledge about evaluation. It does not really matter whether it is acquired through formal training or working experience. Evaluations are conducted for the purpose of improvement; thus it is useful for practitioners to have this knowledge so that the evaluation conducted could meet the function rather than merely meeting the requirements of the interested parties.

5. CONCLUSION

Based on the findings of this study there are six conclusions that can be made. First, there is a problem of scholarly illiteracy among training evaluators in Malaysia. Second, knowledge of conducting evaluation is gained through experience. Third, evaluation focuses on end (after training) evaluation. Fourth, evaluation of training is mainly based on trainees' feedback. Fifth, evaluations done are moderately comprehensive and systematic; and the sixth conclusion is that the practice of evaluation, to a certain extent, is influenced by the background characteristics of the evaluators.

It is obvious that training providers in Malaysia are heavily involved in evaluation and they do have assigned staff responsible for training evaluation. However, the evaluators are not academically trained in the evaluation field and they do not recognize most of the popular evaluation models. This shows that there is a problem of scholarly illiteracy among training evaluators in Malaysia. The findings of this study support the study conducted by Shadish and Epstein (1987). These researchers found that a majority of training evaluators are not familiar with evaluation models and concepts described in the literature, and they were not aware of any evaluation model they used.

This study is significant at the theoretical level because the findings show the low level of evaluation model familiarity among training evaluators. This suggests there is a danger of scholarly illiteracy in evaluation as highlighted by Shadish and Epstein (1987). It also suggests a few possibilities, such as training evaluators do not read sufficiently, or are incompetent, or the theories are irrelevant to practice and the practitioners do not find them useful. The Joint Committee of Standards for Educational Evaluation (1981) requires that the persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that their findings achieve maximum capacity and acceptance. Evaluators are credible to the extent that they exhibit the technical competence, substantive knowledge, experience, integrity, public relations skills, and other characteristics considered necessary by clients and other users of the evaluation report. The empirical findings of this study allow greater insight in to what extent evaluators' background and their understanding of evaluation models actually influence and control the actions of training evaluators. By

taking these into consideration, future evaluation model designers would be more sensitive to the complexity of actual practices of evaluation in the training field. There are very few evaluation models which take training settings into account. Most of the evaluation models are mainly designed for educational and social programs. Though training is partly an educational program, it should be viewed differently because training is conducted within specific setting with the purpose of improving employees' performance to meet the organizational goals. Each industry requires certain types of behavior of its employees for it to function optimally and this set of required behaviors would differ depending on the complexity of the organization's operations. As suggested by Shamsuddin (1995), a specific model should be developed to accommodate for a specific context.

REFERENCES

Adams, Debra, and Catherine Waddle. "Evaluating the Return from Management Development Program: Individual Returns versus Organizational Benefits." *International Journal of Contemporary Hospitality Management* 14, no.1 (2002): 14-20.

Athanasou, James. "A Framework for Evaluating the Effectiveness of Technology-assisted Learning." *Industrial and Commercial Training* 30, no. 3 (1998): 96-103.

Bramley, Peter. "Evaluating Effective Management Learning." *Journal of European Industrial Training* 23, no. 3 (1999): 145-53.

Caffarella, Rosemary S. *Planning Programs for Adult Learners*. San Francisco: Jossey-Bass Publishers, 1994.

Chen, Huey-tsyh, and H. Peter Rossi, eds. *Using Theory to Improve Program and Policy Evaluations*. Westport: Greenwood Press, 1992.

Dixon, Nancy M. "New Routes to Evaluation." *Training & Development Journal* 50, no.5 (1996): 82-5.

Garavaglia, Paul L. "How to Ensure Transfer of Training." *Training & Development Journal* 47, no. 10 (1993): 63-8.

Geertshuis, Susan, Marry Holmes, Harry Geertshuis, David Clancy and Amanda Bristol. "Evaluation of Workplace Learning." *Journal of Workplace Learning* 14, no.1 (2002): 11-8.

Joint Committee on Standards for Educational Evaluation. *Standards*

for Evaluations of Educational Programs, Projects, and Materials. New York: McGraw-Hill, 1981.

Kirkpatrick, Donald L. *Evaluating Training Programs*: *The Four Levels*. San Francisco: Beret-Koehler Publisher, 1994.

Maimunah, Ismail. *Pengembangan: Implikasi ke atas Pembangunan Masyarakat* [*Extension: Implication to Community Development*]. 2nd ed. Kuala Lumpur: Dewan Bahasa dan Pustaka, 1990.

Malaysia.*Eighth Malaysia Plan: 2001-2005*. Kuala Lumpur: Government Printer, 2001.

Mann, Sandi, and T. Ivan Robertson. "What Should Training Evaluations Evaluate?" *Journal of European Industrial Training* 20, no.9 (1996): 14-20.

McClelland, Sam. "A Model for Designing Objective-oriented Training Evaluations." *Industrial and Commercial Training* 26, no. 1 (1994): 3-9.

Phillips, Jack L. *Handbook of Training Evaluation and Measurement Methods*. Houston: Gulf Publishing Company, 1991.

Redshaw, Bernard. "Evaluating Organizational Effectiveness." *Industrial and Commercial Training* 32, no. 7 (2000): 245-48.

Rosti, Robert T., and Frank Shipper. "A Study of the Impact of Training in a Management Development Program based on 360 Feedback." *Journal of Managerial Psychology* 13, no.1/2 (1998): 77-89.

Shadish, William R., and Roberta Epstein. "Patterns of Program Evaluation Practice among Members of the Evaluation Research Society and Evaluation Network." *Evaluation Review* 11, no. 5 (1987): 555-90.

Shamsuddin, Ahmad. "Contextual Factors Associated with Evaluation Practices of Selected Adult and Continuing Education Providers in Malaysia." *Unpublished Ph.D. Dissertation*, University of Georgia, Athens, 1995.

Shelton, Sandra, and George Alliger. "Who's Afraid of Level 4 Evaluation?" *Training & Development Journal* 47, no. 6 (1993): 43-6.

Talbot, Colin. "Evaluation and Validation: A Mixed Approach." *Journal of European Industrial Training* 16, no. 5 (1992): 26-32.