DOCUMENT RESUME

ED 379 350 TM 022 744

AUTHOR Marth, Joseph; And Others
TITLE Evaluating Organizational Training Programs: Alternatives and Criteria for Selection.
PUB DATE Oct 94
NOTE 12p.; Paper presented at the Annual Meeting of the Midwestern Educational Research Association (Chicago, IL, October 1994).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Criteria; *Decision Making; *Evaluation Methods; Matrices; Objectives; *Organizations (Groups); *Program Evaluation; *Selection; Surveys; *Training
IDENTIFIERS *Utility Analysis

ABSTRACT

Managers and program sponsors are often unaware of possible alternatives to costly training evaluation procedures and do not have criteria for selecting alternatives. What is needed is an understanding of the various levels of evaluating training programs, feasible alternatives, and decision criteria for choosing the right system. It is proposed that Kirkpatrick's four levels of training evaluation, plus three enhancements be used: (1) Level 2 enhancement, perceptions of objective mastery; (2) Level 3 enhancement, perceptions of objective transfer; and (3) Level 4 enhancement, utility theory estimates. The first two alternatives involve using surveys and existing instructional objectives. The last alternative involves utility calculations based on the Schmidt, Hunter, and Pearlman (1982) model. Decisions about implementing training evaluation and the degree of sophistication should be based on the criteria of feasibility, propriety, utility, and accuracy. By mapping these criteria against the four-plus levels of evaluation, a decision matrix is obtained, and organizational trainers and educators are able to determine the appropriate degree of evaluation of their training programs. One figure presents the decision matrix. (Contains 15 references.) (Author/SLD)

Evaluating Organizational Training Programs:

Alternatives and Criteria for Selection

Joseph Marth, Ph.D.
Bluefield College

Burton Hancock, Ph.D.
Ernst & Young

and

Bonnie Cirignano, Ph.D.
The University of Akron

Presented at the Midwestern Educational Research Association Conference, Chicago
October 1994

Running Head: EVALUATING TRAINING

Abstract

Regrettably, managers and program sponsors are often unaware of possible alternatives to costly training evaluation procedures and do not have criteria for their selection. What is needed, therefore, is an understanding of the various levels of evaluating training programs, feasible alternatives, and decision criteria for choosing the right system. We propose using Kirkpatrick's four levels of training evaluation, plus three enhancements: Level II *Enhancement*: Perceptions of Objective Mastery; Level III *Enhancement*: Perceptions of Objective Transfer; and Level IV *Enhancement*: Utility Theory Estimates. The first two alternatives involve using surveys and existing instructional objectives. The last alternative involves utility calculations based on the Schmidt, Hunter, and Pearlman (1982) model.

Decisions about implementing training evaluation and the degree sophistication should be based on the following four criteria: (1) Feasibility; (2) Propriety; (3) Utility; and (4) Accuracy. By mapping these criteria against the 4plus levels of evaluation, a decision matrix is obtained and organizational trainers/educators are able to determine the appropriate degree of evaluation of their training programs.

Suppose for the moment that your organization spent millions of dollars annually on training programs, and that your success in business depended on having highly trained employees. No doubt you would consider carefully evaluating the programs' effectiveness and efficiency. We know that this only makes sense – formal evaluation is the "quality control process" whereby we can assure the validity of our training programs.

In reality, however, evaluation often never takes place. According to Carnevale and Schulz (1990), many training professionals contend "that accounting for training (through measurement and evaluation) takes too much time or is too costly" (p. s–2). Training is typically taken on faith that it is valid and meetings its intended objectives. As Carnevale and Schulz (1990) comment:

> [t]he fact that fewer than half of America's training programs are formally evaluated indicates implicit managerial trust that, somehow or other, training facilitates attainment of organizational goals. (p. s–2)

Many organizations believe that they can implicitly tell whether their training is working. And surprisingly, they probably can – in a very crude sense. Obviously, if you're losing consulting engagements because your consultants are not properly trained to deliver effective business solutions, you'll know it. Unfortunately, without a formal evaluation system, the organization only has a sense that its training programs aren't working properly. They will not necessarily know which programs aren't working well, where the programs are failing, and the root causes of these failures. Further, without evaluations, one can neither efficiently pinpoint changes nor avoid subjective revisions.

One can theorize additional reasons why organizations are reluctant to institute evaluation systems. For example, one reason for not instituting

program evaluation we typically encounter relates to the company's culture. Many top managers believe that testing, the normal means to evaluate learning, is countercultural. In other words, testing is reserved for academics and this is the business world; our people wouldn't tolerate taking tests.

A second reason has to do with the impact of a poor evaluation. If you are the program owner/manager and you determine the program is not effective, one consequence may be that your performance would be viewed negatively. Furthermore, poor evaluations increase the risks of reduced funding, loss of program sponsorship, and the decision to use outside training vendors.

Third, the nature of the subject matter, organization/management development in our case, is in a continuous state of flux and change. This instability and the constant evolution of programs precludes one from sometimes developing relatively expensive evaluation systems which would become outdated in a matter of weeks.

Finally, limited resources (both human and financial) are usually allocated to developing and implementing new and existing training programs. Consequently, evaluation of existing programs is given considerably lower priority than rapidly putting out new training programs. Unfortunately, without evaluation, we gain little information about whether the new or existing programs are meeting the needs of the organization.

While the costs of evaluating training programs can be high, the consequences of not evaluating training programs can be even costlier to organizations. Failures to demonstrate the validity of training programs has resulted in the untimely demise of valid training programs – training programs that brought about important organizational changes and/or future revenues. Further, the costs associated with ineffective or counter effective training

programs can be deadly to organizations attempting to compete on the basis of cost, quality, and timeliness in a global economy.

A Systematic Approach to Training

Before we discuss the alternatives for evaluating training and their selection criteria, it is first important to provide a context for training evaluation by providing an overview of the training process. Goldstein's (1993) systems approach has training beginning with a needs assessment stage where one conducts three analyses to determine instructional need: (1) organizational analysis; (2) task and KSA (knowledge, skills, abilities) analysis; and (3) person analysis. The second stage to the training process entails converting the data obtained from the needs assessment to the actual development of the training program. This stage requires the development of terminal instructional objectives and the selection of appropriate instructional delivery modalities. The third stage is the actual implementation and delivery of the training program. The final stage, the evaluation stage, includes two sub-stages: (1) *formative evaluation* which evaluates the program throughout the entire process; and (2) *summative evaluation* which assesses the program at its completion and/or after a later period of time (Scriven, 1967).

Alternatives to Evaluation: Kirkpatrick's Four Levels Plus 3 Enhancements

We propose that the evaluation of training programs should be based upon Kirkpatrick's (1959; 1960) four levels, with the addition of several enhancements. These 4plus levels can be progressively implemented depending on the resources and skills of the training professionals and training department. Finally, we believe there are four important decision criteria which can be used for selecting the appropriate training evaluation approach.

Our 4plus phase system of evaluation is built upon the pioneering work of Kirkpatrick (1959; 1960), who envisioned the summative evaluation of training entailing four levels of sophistication: Level I: Reaction; Level II: Learning; Level III: Transfer; Level IV: Organizational Impact. To these four levels we propose adding three enhancements: Level II *Enhancement*: Perceptions of Objective Mastery; Level III *Enhancement*: Perceptions of Objective Transfer; and Level IV *Enhancement*: Utility Theory Estimates. The first two enhancements involve using surveys and existing instructional objectives. The last enhancement involves utility calculations based on the Schmidt, Hunter, and Pearlman (1982) model.

Level I measures participants' reactions to the training program. It is the least sophisticated of the levels of evaluation and has been nicknamed the "smiles" assessment. The intent of this form of evaluation is obtain an immediate indicator of participant's impressions of the learning experience.

Level II measures participants' learning – did the participants learn what they were intended to learn. Typically this is accomplished through the use of objective testing procedures. Ideally, participants knowledge should by measured before training (pre-testing) and after training (post-testing), with positive gain scores indicating the significance of their learning.

Level II *Enhancement* measures learning by surveying program participants, immediately upon completion of a module or program, regarding their mastery of the learning objectives. Participants are asked to rate their degree of agreement (1=Strongly Disagree; 2=Disagree; 3=Neutral; 4=Agree; 5=Strongly Agree) with statements indicating mastery of the program's terminal learning objectives as a result of the training experience.

Level III measures the extent to which learning/training transfers to the job. Ideally, participants' behavior back on the job is objectively observed and recorded to determine if the training experience has transferred.

Level III *Enhancement* measures transfer of training by taking the program's terminal learning objectives and then surveying participants (and their managers) about which program objectives transferred/applied to the job. This should be conducted after a set period of time (e.g., three months). This is also a good time to inquire about areas in which participants should have received training in, but did not. This procedure closely dovetails the program's needs assessment phase.

Level IV measures the extent to which the training experience has had an impact on organizational goals or objectives. Examples include the reduction of accidents, reduced scrap, cost savings, increased customer satisfaction, reduced grievances, and etc. Typically this requires researching specific organizational components and attempting to isolate the effects of the training program on organizational goals.

Level IV *Enhancement* involves utility calculations based on the Schmidt, Hunter, and Pearlman (1982) model. Utility theory originates from recent work in personnel psychology whereby organizations can weigh the financial costs and benefits of human resource programs, such as selection and training (Becker, 1989). The utility of a training program is the extent to which the training experience improves the quality of those participating beyond that which would have occurred if that program had not been provided. Quality can include the full range or organizational objectives stated earlier, but usually refers to the dollar payoff of the training program for the o ganization. Cascio (1987) recommends the following formula developed by Schmidt, Hunter, and Pearlman

(1982) be used for evaluating training programs: $\Delta U = (T)(N)(d_t)(SD_y) - (N)(C)$, where

ΔU = the dollar value of a training program

T = number of years duration of the training on the performance area of interest

N = the number of persons trained

d_t = the true difference in performance between the average trained and the average untrained employee in units of standard deviation. [Note: This is an effect size estimate.]

SD_y = the standard deviation of performance in dollars of the untrained group.

C = the cost of training per individual.

Finally, it should be noted that Mathieu and Leonard (1987) have extended the work of Boudreau (1983) and provide a modified utility formula which incorporates the influence of employee flows and economic factors over time.

Criteria for Conducting Training Assessment

Further examination of these 4plus levels reveals that each has its associated sophistication, benefits, and limitations. We believe that decisions about implementing training evaluation should be based on the following four criteria: (1) Feasibility; (2) Propriety; (3) Utility; and (4) Accuracy. Feasibility refers to whether the evaluation approach is realistically (logistically and economically) viable. Propriety relates to whether the evaluation approach is legally, ethically, and politically and culturally viable. Utility refers to whether the evaluation approach serves the needs of the various organizational constituencies. And finally, accuracy relates to whether the approach is technically sound, systematic, valid, and reliable.

By mapping these criteria against the 4plus levels of evaluation, the following decision matrix is obtained:

	Feasibility (High= 3, Med=2., Low=1)	**Propriety** (High= 3, Med=2., Low=1)	**Utility** (High= 3, Med=2., Low=1)	**Accuracy** (High= 3, Med=2., Low=1)	Total
Level I					
Level II					
Level II Enh.					
Level III					
Level III Enh.					
Level IV					
Level IV Enh.					

Organizational trainers/educators may wish to use the matrix to determine the viability of the various approaches to evaluating their training programs. The matrix will likely indicate which training evaluation approaches are most appropriate for the various program parameters and organizational constituencies.

In conclusion, the need for training evaluation is greater today than ever. As Carnevale and Schulz (1990) state,

> [h]uman resource development (HRD) professionals reluctant to account for training need to reorient their thinking to face the business realities of the nineties. Instead of deciding *whether* to measure and evaluate training's results, they must decide *how* to determine its costs and benefits. (p. s–2)

We hope that this paper will assist in this important endeavor.

References

Becker, B. E. (1989). The influence of labor markets on human resources utility estimnates. Personnel Psychology, 42: 531-546.

Boudreau, J. W. (1983). Economic Considerations in estimating the utility of human resource productivity improvement programs. Personnel Psychology, 36: 551-576

Carnevale, A. P. & Schulz, E. R. (1990). Return on Investment: Accounting for Training. Training & Development Journal (Supplement, July, 1990).

Cascio, W. E. (1987). Applied Psychology in Personnel Management (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Goldstein, I. L. (1993). Training in Organizations (3rd ed.). Belmont, CA: Wadsworth.

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. Journal of the American Society of Training Directors, 13: 3-26.

Kirkpatrick, D. L. (1960). Techniques for evaluating training programs. Journal of the American Society of Training Directors, 14: 13-32.

Mathieu, J. E. & Leonard, R. L. (1987). Applying utility concepts to a training program in supervisory skills: A time-based approach. Academy of Management Journal, 30 (2): 316-335.

Schmidt, F. L., Hunter, J.E., & Pearlman, K. (1982). Assessing the economic impact of personnel programs on workforce productivity. Personnel Psychology, 35: 333-347.

Scriven, M. (1967). The methodology of evaluation. In Perspectives of Curriculum Evaluation. American Educational Research Association Monograph, 1.