5

Evaluation Models, Approaches, and Designs

BACKGROUND

This section includes activities that address

- Understanding and selecting evaluation models and approaches
- Understanding and selecting evaluation designs

The following information is provided as a brief introduction to the topics covered in these activities.

EVALUATION MODELS AND APPROACHES

The following models and approaches are frequently mentioned in the evaluation literature.

Behavioral Objectives Approach. This approach focuses on the degree to which the objectives of a program, product, or process have been achieved. The major question guiding this kind of evaluation is, "Is the program, product, or process achieving its objectives?"

The Four-Level Model. This approach is most often used to evaluate training and development programs (Kirkpatrick, 1994). It focuses on four levels of training outcomes: reactions, learning, behavior, and results. The major question guiding this kind of evaluation is, "What impact did the training

101

have on participants in terms of their reactions, learning, behavior, and organizational results?"

Responsive Evaluation. This approach calls for evaluators to be responsive to the information needs of various audiences or stakeholders. The major question guiding this kind of evaluation is, "What does the program look like to different people?"

Goal-Free Evaluation. This approach focuses on the actual outcomes rather than the intended outcomes of a program. Thus, the evaluator has minimal contact with the program managers and staff and is unaware of the program's stated goals and objectives. The major question addressed in this kind of evaluation is, "What are all the effects of the program, including any side effects?"

Adversary/Judicial Approaches. These approaches adapt the legal paradigm to program evaluation. Thus, two teams of evaluators representing two views of the program's effects argue their cases based on the evidence (data) collected. Then, a judge or a panel of judges decides which side has made a better case and makes a ruling. The question this type of evaluation addresses is, "What are the arguments for and against the program?"

Consumer-Oriented Approaches. The emphasis of this approach is to help consumers choose among competing programs or products. *Consumer Reports* provides an example of this type of evaluation. The major question addressed by this evaluation is, "Would an educated consumer choose this program or product?"

Expertise/Accreditation Approaches. The accreditation model relies on expert opinion to determine the quality of programs. The purpose is to provide professional judgments of quality. The question addressed in this kind of evaluation is, "How would professionals rate this program?"

Utilization-Focused Evaluation. According to Patton (1997), "utilization-focused program evaluation is evaluation done for and with specific, intended primary users for specific, intended uses" (p. 23). As such, it assumes that stakeholders will have a high degree of involvement in many, if not all, phases of the evaluation. The major question being addressed is, "What are the information needs of stakeholders, and how will they use the findings?"

Participatory/Collaborative Evaluation. The emphasis of participatory/collaborative forms of evaluation is engaging stakeholders in the evaluation process, so they may better understand evaluation and the program being evaluated and ultimately use the evaluation findings for decision-making

purposes. As with utilization-focused evaluation, the major focusing question is, "What are the information needs of those closest to the program?"

Empowerment Evaluation. This approach, as defined by Fetterman (2001), is the "use of evaluation concepts, techniques, and findings to foster improvement and self-determination" (p. 3). The major question characterizing this approach is, "What are the information needs to foster improvement and self-determination?"

Organizational Learning. Some evaluators envision evaluation as a catalyst for learning in the workplace (Preskill & Torres, 1999). Thus, evaluation can be viewed as a social activity in which evaluation issues are constructed by and acted on by organization members. This approach views evaluation as ongoing and integrated into all work practices. The major question in this case is, "What are the information and learning needs of individuals, teams, and the organization in general?"

Theory-Driven Evaluation. This approach to evaluation focuses on theoretical rather than methodological issues. The basic idea is to use the "program's rationale or theory as the basis of an evaluation to understand the program's development and impact" (Smith, 1994, p. 83). By developing a plausible model of how the program is supposed to work, the evaluator can consider social science theories related to the program as well as program resources, activities, processes, and outcomes and assumptions (Bickman, 1987). The major focusing questions here are, "How is the program supposed to work? What are the assumptions underlying the program's development and implementation?"

Success Case Method. This approach to evaluation focuses on the practicalities of defining successful outcomes and success cases (Brinkerhoff, 2003) and uses some of the processes from theory-driven evaluation to determine the linkages, which may take the form of a logic model, an impact model, or a results map. Evaluators using this approach gather stories within the organization to determine what is happening and what is being achieved. The major question this approach asks is, "What is really happening?"

EVALUATION DESIGNS

Evaluation designs that collect quantitative data fall into one of three categories:

1. Preexperimental

2. Quasi-experimental

3. True experimental designs

The following are brief descriptions of the most commonly used evaluation (and research) designs.

One-Shot Design. In using this design, the evaluator gathers data following an intervention or program. For example, a survey of participants might be administered after they complete a workshop.

Retrospective Pretest. As with the one-shot design, the evaluator collects data at one time but asks for recall of behavior or conditions prior to, as well as after, the intervention or program.

One-Group Pretest-Posttest Design. The evaluator gathers data prior to and following the intervention or program being evaluated.

Time Series Design. The evaluator gathers data prior to, during, and after the implementation of an intervention or program.

Pretest-Posttest Control-Group Design. The evaluator gathers data on two separate groups prior to and following an intervention or program. One group, typically called the experimental or treatment group, receives the intervention. The other group, called the control group, does not receive the intervention.

Posttest-Only Control-Group Design. The evaluator collects data from two separate groups following an intervention or program. One group, typically called the experimental or treatment group, receives the intervention or program, while the other group, typically called the control group, does not receive the intervention. Data are collected from both of these groups only after the intervention.

Case Study Design. When evaluations are conducted for the purpose of understanding the program's context, participants' perspectives, the inner dynamics of situations, and questions related to participants' experiences, and where generalization is not a goal, a case study design, with an emphasis on the collection of qualitative data, might be most appropriate. Case studies involve in-depth descriptive data collection and analysis of individuals, groups, systems, processes, or organizations. In particular, the case study design is most useful when you want to answer *how* and *why* questions and when there is a need to understand the particulars, uniqueness, and diversity of the case.

RETURN-ON-INVESTMENT DESIGNS

Many evaluations, particularly those undertaken within an organizational setting, focus on financial aspects of a program. Typically in such evaluations,

the questions involve a program's "worth." Four primary approaches include cost analysis, cost-benefit analysis, cost-effectiveness analysis, and return on investment (ROI).

Cost analysis involves determining all of the costs associated with a program or an intervention. These need to include trainee costs (time, travel, and productivity loss), instructor or facilitator costs, materials costs, facilities costs, as well as development costs. Typically, a cost analysis is undertaken to decide among two or more different alternatives for a program, such as comparing the costs for in-class delivery versus online delivery.

Cost analyses examine only costs. A cost-effectiveness analysis determines the costs as well as the direct outcomes or results of the program. As with cost analyses, the costs are measured in dollars or some other monetary unit. The effectiveness measure may include such things as reduced errors or accidents, improved customer satisfaction, and new skills. The decision maker must decide whether the costs justify the outcomes.

A cost-benefit analysis transforms the effects or results of a program into dollars or some other monetary unit. Then the costs (also calculated in monetary terms) can be compared to the benefits. As an example, let us assume that a modification in the production system is estimated to reduce errors by 10%. Given that production errors cost the company $1,000,000 last year, the new system should save the company $100,000 in the first year and the succeeding year. Assuming that the modification would cost $100,000 and the benefits would last for 3 years, we can calculate the benefit/cost ratio as follows:

Benefit/cost ratio = Program benefits/program costs

Benefit/cost ratio = $300,000/$100,000

Benefit/cost ratio = 3:1

This means that for each dollar spent, the organization would realize three dollars of benefits.

The ROI calculation is often requested by executives. Using the previous example, the formula is as follows:

ROI = [Net program benefits/Program costs] × 100%

ROI = [(Program benefits – Program costs)/Program costs] × 100%

ROI = [($300,000 – $100,000)/$100,000] × 100%

ROI = [$200,000/$100,000] × 100%

ROI = 2 × 100%

ROI = 200%

This means that the costs were recovered, and an additional 200% of the costs were returned as benefits.

RESOURCES

Alkin, M. C. (Ed.). (2004). *Evaluation roots: Tracing theorists' views and influences.* Thousand Oaks, CA: Sage.

Bickman, L. (1987). The function of program theory. In P. J. Rogers, T. A. Haccsi, A. Petrosino, & T. A. Huebner (Eds.), *Using program theory in education* (New Directions for Program Evaluation, Vol. 33, pp. 5-18). San Francisco: Jossey Bass.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: Handbook I: Cognitive domain.* New York: David McKay.

Brigham, E. F., Gapenski, L. C., & Ehrhart, M. C. (1998). *Financial management: Theory and practice* (9th ed.). New York: Thomson.

Brinkerhoff, R. O. (2003). *The success case method: Find out quickly what's working and what's not.* San Francisco: Berrett-Koehler.

Chen, H. T. (1990). *Theory-driven evaluations.* Newbury Park, CA: Sage.

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation* (New Directions for Evaluation, Vol. 80, pp. 5-23). San Francisco: Jossey-Bass.

Fetterman, D. M. (2001). *Foundations of empowerment evaluation.* Thousand Oaks, CA: Sage.

House, E. R. (1993). *Professional evaluation: Social impact and political consequences.* Thousand Oaks, CA: Sage.

Kee, J. E. (1994). Benefit-cost analysis in program evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 456-488). San Francisco: Jossey-Bass.

Kirkpatrick, D. (1994). *Evaluating training programs: The four levels.* San Francisco: Berrett-Koehler.

Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis: Methods and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Mager, R. F. (1962). *Preparing instructional objectives.* Palo Alto, CA: Fearon Press.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs.* San Francisco: Jossey-Bass.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text.* Thousand Oaks, CA: Sage.

Phillips, J. J. (1997). *Return on investment in training and development programs.* Houston, TX: Gulf Publishing.

Preskill, H., & Torres, R. T. (1999). *Evaluative inquiry for learning in organizations.* Thousand Oaks, CA: Sage.

Russ-Eft, D., & Preskill, H. (2001). *Evaluation in organizations: A systematic approach to learning, performance, and change.* Boston: Perseus.

Scriven, M. (1973). Goal-free evaluation. In E. R. House (Ed.), *School evaluation* (pp. 319-328). Berkeley, CA: McCutchan.

Scriven, M. (1994). Product evaluation—The state of the art. *Evaluation Practice, 15*(1), 45-62.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1995). *Foundations of program evaluation: Theories of practice.* Thousand Oaks, CA: Sage.

Smith, N. L. (1994). Clarifying and expanding the application of program theory-driven evaluations. *Evaluation Practice, 15*(1), 83-87.

Stake, R. E. (1995). *The art of case study research.* Thousand Oaks, CA: Sage.

Stake, R. E. (2004). *Standards-based and responsive evaluation.* Thousand Oaks, CA: Sage.

Stufflebeam, D. L. (Ed.). (2001). *Evaluation models* (New Directions for Evaluation, Vol. 89). San Francisco: Jossey-Bass.

Swanson, R. A., & Holton, E. F., III (1999). *Results: How to assess performance, learning, and perceptions in organizations.* San Francisco: Berrett-Koehler.

Wolf, R. L. (1975). Trial by jury: A new evaluation method. *Phi Delta Kappan, 57,* 185-187.

Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). *Program evaluation: Alternative approaches and practice guidelines* (2nd ed.). New York: Longman.

Yin, R. K. (2002). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.

Activity 20

Determining When and Where to Use Various Evaluation Models and Approaches

Overview

This activity provides participants with an understanding of various evaluation models and approaches and how they can be used.

Instructional Objectives

Participants will

- Describe the conditions under which certain evaluation models or approaches may be most effective or appropriate
- Discuss the implications of using various evaluation models and approaches for an evaluation study
- Discuss when and how one chooses to use a particular evaluation model or approach

Number of Participants

- Minimum number of participants: 3
- Maximum number of participants: unlimited when participants are in groups of 3 to 5

Time Estimate: 45 to 60 minutes

In addition to providing the necessary background information on various evaluation models and approaches, this activity requires approximately 45 to 60 minutes, depending on the number of participants (or groups) and the time available for discussion.

Materials Needed

- Pens/pencils
- Flipchart, markers, tape
- Handout "Evaluation Models and Approaches"

Instruction Method

Small-group work

Procedures

Facilitator's tasks:

- Ask participants to get into groups of three to five people.
- Depending on the number of groups, distributes one or two different handouts (models and approaches) to each group.
- Instruct participants, as a group, to complete the handout.
- Invite groups to share their ideas with the larger group. Ask other participants to add their ideas if they worked on this model or approach.
- Debrief the activity with the following questions:
  – Which models or approaches seem similar or compatible? In what ways are they similar or compatible?
  – Which models or approaches have different orientations? How might these differences manifest themselves in an evaluation?
  – Which of the models or approaches would fit within the context of the organization or organizations with which you typically work?
  – How do you think one decides which models and approaches to use for any one evaluation? What criteria would you use to determine the most appropriate model and approach for a given evaluation context?

Evaluation Models and Approaches

Handout for Activity 20

Behavioral Objectives

This approach focuses on the degree to which the objectives of a program, product, or process have been achieved. The major question guiding this kind of evaluation is, "Is the program, product, or process achieving its objectives?"

What are some examples or situations in which you would use this approach?	*What conditions need to exist to use this approach?*	*What are some limitations of this approach?*

Evaluation Models and Approaches

Handout for Activity 20

The Four-Level Model

This approach is most often used to evaluate training and development programs (Kirkpatrick, 1994). It focuses on four levels of training outcomes: reactions, learning, behavior, and results. The major question guiding this kind of evaluation is, "What impact did the training have on participants in terms of their reactions, learning, behavior, and organizational results?"

What are some examples or situations in which you would use this approach?	*What conditions need to exist to use this approach?*	*What are some limitations of this approach?*

Reference

Kirkpatrick, D. (1994). *Evaluating training programs: The four levels.* San Francisco: Berrett-Koehler.

Evaluation Models and Approaches

Handout for Activity 20

Responsive Evaluation

This approach calls for evaluators to be responsive to the information needs of various audiences or stakeholders. The major question guiding this kind of evaluation is, "What does the program look like to different people?"

What are some examples or situations in which you would use this approach?	*What conditions need to exist to use this approach?*	*What are some limitations of this approach?*

Evaluation Models and Approaches

Handout for Activity 20

Goal-Free Evaluation

This approach focuses on the actual outcomes rather than the intended outcomes of a program. Thus, the evaluator has minimal contact with the program managers and staff and is unaware of the program's stated goals and objectives. The major question addressed in this kind of evaluation is, "What are all the effects of the program, including any side effects?"

What are some examples or situations in which you would use this approach?	*What conditions need to exist to use this approach?*	*What are some limitations of this approach?*

Evaluation Models and Approaches

Handout for Activity 20

Adversary/Judicial

These approaches adapt the legal paradigm to program evaluation. Thus, two teams of evaluators representing two views of the program's effects argue their case based on the evidence (data) collected. Then, a judge or a panel of judges decides which side made a better case and makes a ruling. The question this type of evaluation addresses is, "What are the arguments for and against the program?"

What are some examples or situations in which you would use this approach?	*What conditions need to exist to use this approach?*	*What are some limitations of this approach?*

Evaluation Models and Approaches

Handout for Activity 20

Consumer-Oriented

The emphasis in this approach is to help consumers choose among competing programs or products. *Consumer Reports* provides an example of this type of evaluation. The major question addressed by this evaluation is, "Would an educated consumer choose this program or product?"

What are some examples or situations in which you would use this approach?	*What conditions need to exist to use this approach?*	*What are some limitations of this approach?*

Evaluation Models and Approaches

Handout for Activity 20

Utilization-Focused

According to Patton (1997), "utilization-focused program evaluation is evaluation done for and with specific, intended primary users for specific, intended uses" (p. 23). As such, it assumes that stakeholders will have a high degree of involvement in many, if not all, phases of the evaluation. The major question being addressed is, "What are the information needs of stakeholders, and how will they use the findings?"

What are some examples or situations in which you would use this approach?	*What conditions need to exist to use this approach?*	*What are some limitations of this approach?*

Reference

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text.* Thousand Oaks, CA: Sage.

116

Evaluation Models and Approaches

Handout for Activity 20

Expertise/Accreditation

The accreditation model relies on expert opinion to determine the quality of programs. The purpose is to provide professional judgments of quality. The question addressed in this kind of evaluation is, "How would professionals rate this program?"

What are some examples or situations in which you would use this approach?	*What conditions need to exist to use this approach?*	*What are some limitations of this approach?*

Evaluation Models and Approaches

Handout for Activity 20

Participatory/Collaborative

The emphasis of participatory/collaborative forms of evaluation is engaging stakeholders in the evaluation process, so they may better understand evaluation and the program being evaluated and ultimately use the evaluation findings for decision-making purposes. As with utilization-focused evaluation, the major focusing question is, "What are the information needs of those closest to the program?"

What are some examples or situations in which you would use this approach?	*What conditions need to exist to use this approach?*	*What are some limitations of this approach?*

Evaluation Models and Approaches

Handout for Activity 20

Empowerment

This approach, as defined by Fetterman (2001), is the "use of evaluation concepts, techniques, and findings to foster improvement and self-determination" (p. 3). The major question characterizing this approach is, "What are the information needs to foster improvement and self-determination?"

What are some examples or situations in which you would use this approach?	*What conditions need to exist to use this approach?*	*What are some limitations of this approach?*

Reference

Fetterman, D. M. (2001). *Foundations of empowerment evaluation.* Thousand Oaks, CA: Sage.

Evaluation Models and Approaches

Handout for Activity 20

Organizational Learning

Some evaluators envision evaluation as a catalyst for learning in the workplace (Preskill & Torres, 1999). Thus, evaluation can be viewed as a social activity in which evaluation issues are constructed by and acted on by organizational members. This approach views evaluation as ongoing and integrated into all work practices. The major question in this case is, "What are the information and learning needs of individuals, teams, and the organization in general?"

What are some examples or situations in which you would use this approach?	*What conditions need to exist to use this approach?*	*What are some limitations of this approach?*

Reference

Preskill, H., & Torres, R. T. (1999). *Evaluative inquiry for learning in organizations.* Thousand Oaks, CA: Sage.

120

Evaluation Models and Approaches

Handout for Activity 20

Theory-Driven

This approach to evaluation focuses on theoretical rather than methodological issues. The basic idea is to use the "program's rationale or theory as the basis of an evaluation to understand the program's development and impact" (Smith, 1994, p. 83). By developing a plausible model of how the program is supposed to work, the evaluator can consider social science theories related to the program as well as program resources, activities, processes, and outcomes and assumptions (Bickman, 1987). The major focusing questions of this approach are, "How is the program supposed to work? What are the assumptions underlying the program's development and implementation?"

What are some examples or situations in which you would use this approach?	*What conditions need to exist to use this approach?*	*What are some limitations of this approach?*

References

Bickman, L. (1987). The function of program theory. In P. J. Rogers, T. A. Haccsi, A. Petrosino, & T. A. Huebner (Eds.), *Using program theory in education* (New Directions for Program Evaluation, Vol. 33, pp. 5-18). San Francisco: Jossey Bass.

Smith, N. L. (1994). Clarifying and expanding the application of program theory-driven evaluations. *Evaluation Practice, 15*(1), 83-87.

Evaluation Models and Approaches

Handout for Activity 20

Success Case Method

This approach to evaluation focuses on the practicalities of defining successful outcomes and success cases (Brinkerhoff, 2003) and uses some of the processes from theory-driven evaluation to determine the linkages, which may take the form of a logic model, an impact model, or a results map. Evaluators using this approach then gather success stories within the organization to determine what is happening and what is being achieved. The major question this approach asks is, "What is really happening?"

What aspects of the training will you evaluate?	What variables will you focus on?	What are some potential limitations of the evaluation and its findings?

Reference

Brinkerhoff, R. O. (2003). *The success case method: Find out quickly what's working and what's not.* San Francisco: Berrett-Koehler.

Activity 21

Recommending an Evaluation Approach

Overview

This activity asks participants to consider several evaluation approaches and to choose one or more that would serve the stakeholders' information needs.

Instructional Objectives

Participants will

- Learn about and discuss various approaches to conducting an evaluation and the relative merits of each one
- Read a case scenario and choose one or more approaches that address the questions posed in the case
- Present the reasons for selecting a particular evaluation approach

Number of Participants

- Minimum number of participants: 6
- Maximum number of participants: 25

Time Estimate: 45 to 90 minutes

In addition to providing the necessary background information about different approaches to conducting an evaluation, this activity requires approximately 45 to 90 minutes, depending on the number of participants (or groups) and the time available for discussion.

Materials Needed

- Pens/pencils
- Flipchart, markers, tape

- Handout "Program Evaluation Approaches"
- Handout "Evaluating a Sales Training Program"
- Handout "Evaluating a Great Books Reading Program"

Instructional Method

Case scenario

Procedures

Facilitator's tasks:

- Decide whether to have participants work on one of the case scenarios, let participants choose one of the scenarios in their small groups, or divide the group in half and assign one scenario to each group.
- Ask participants to get into groups of three to five people.
- Provide participants with the handout "Program Evaluation Approaches" and either or both of the case scenarios: "Evaluating a Sales Training Program" and "Evaluating a Great Books Program."
- Ask groups to read their case scenarios and answer the questions posed on the handout.
- Instruct groups that they will have 3 minutes to make their presentations to the vice president of sales and the director of marketing.
- Invite groups to make their presentations.
- Debrief the activity by asking the following questions:
  – What did you learn from this activity?
  – What other information would have helped you choose your approach?
  – To what extent do you think the choice of approach affects the outcomes of an evaluation?
  – How does one choose an evaluation approach? What criteria drove your choices?

Program Evaluation Approaches

Handout for Activity 21

Behavioral Objectives Approach. This approach focuses on the degree to which the objectives of a program, product, or process have been achieved. The major question guiding this kind of evaluation is, "Is the program, product, or process achieving its objectives?"

The Four-Level Model. This approach is most often used to evaluate training and development programs (Kirkpatrick, 1994). It focuses on four levels of training outcomes: reactions, learning, behavior, and results. The major question guiding this kind of evaluation is, "What impact did the training have on participants in terms of their reactions, learning, behavior, and organizational results?"

Responsive Evaluation. This approach calls for evaluators to be responsive to the information needs of various audiences or stakeholders. The major question guiding this kind of evaluation is, "What does the program look like to different people?"

Goal-Free Evaluation. This approach focuses on the actual outcomes rather than the intended outcomes of a program. Thus, the evaluator has minimal contact with the program managers and staff and is unaware of the program's stated goals and objectives. The major question addressed in this kind of evaluation is, "What are all the effects of the program, including any side effects?"

Adversary/Judicial Approaches. These approaches adapt the legal paradigm to program evaluation. Thus, two teams of evaluators representing two views of the program's effects argue their case based on the evidence (data) collected. Then, a judge or a panel of judges decides which side made a better case and makes a ruling. The question this type of evaluation addresses is, "What are the arguments for and against the program?"

Consumer-Oriented Approaches. The emphasis in this approach is to help consumers choose among competing programs or products. *Consumer Reports* provides an example of this type of evaluation. The major question addressed by this evaluation is, "Would an educated consumer choose this program or product?"

Expertise/Accreditation Approaches. The accreditation model relies on expert opinion to determine the quality of programs. The purpose is to provide professional

SOURCE: From Russ-Eft and Preskill (2001).

125

judgments of quality. The question addressed in this kind of evaluation is, "How would professionals rate this program?"

Utilization-Focused Evaluation. According to Patton (1997), "utilization-focused program evaluation is evaluation done for and with specific, intended primary users for specific, intended uses" (p. 23). As such, it assumes that stakeholders will have a high degree of involvement in many, if not all, phases of the evaluation. The major question being addressed is, "What are the information needs of stakeholders, and how will they use the findings?"

Participatory/Collaborative Evaluation. The emphasis of participatory/collaborative forms of evaluation is engaging stakeholders in the evaluation process, so they may better understand evaluation and the program being evaluated and ultimately use the evaluation findings for decision-making purposes. As with utilization-focused evaluation, the major focusing question is, "What are the information needs of those closest to the program?"

Empowerment Evaluation. This approach, as defined by Fetterman (2001), is the "use of evaluation concepts, techniques, and findings to foster improvement and self-determination. It is attentive to empowering processes and outcomes" (p. 3). The major question characterizing this approach is, "What are the information needs to foster improvement and self-determination?"

Organizational Learning. Some evaluators envision evaluation as a catalyst for learning in the workplace (Preskill & Torres, 1999). Thus, evaluation can be viewed as a social activity in which evaluation issues are constructed by and acted on by organizational members. This approach views evaluation as ongoing and integrated into all work practices. The major question in this case is, "What are the information and learning needs of individuals, teams, and the organization in general?"

Theory-Driven Evaluation. This approach to evaluation focuses on theoretical rather than methodological issues. The basic idea is to use the "program's rationale or theory as the basis of an evaluation to understand the program's development and impact" (Smith, 1994, p. 83). By developing a plausible model of how the program is supposed to work, the evaluator can consider social science theories related to the program as well as program resources, activities, processes, and outcomes and assumptions (Bickman, 1987). The major focusing questions with this approach are, "How is the program supposed to work? What are the assumptions underlying the program's development and implementation?"

Success Case Method. This approach to evaluation focuses on the practicalities of defining successful outcomes and success cases (Brinkerhoff, 2003). The approach uses some of the processes from theory-driven evaluation to determine the linkages, which may take the form of a logic model, an impact model, or a results map. Evaluators using this approach then gather stories within the organization to

determine what is happening and what is being achieved. The major question this approach asks is, "What is really happening?"

References

Bickman, L. (1987). The function of program theory. In P. J. Rogers, T. A. Haccsi, A. Petrosino, & T. A. Huebner (Eds.), *Using program theory in education* (New Directions for Program Evaluation, Vol. 33, pp. 5-18). San Francisco: Jossey Bass.

Brinkerhoff, R. O. (2003). *The success case method: Find out quickly what's working and what's not.* San Francisco: Berrett-Koehler.

Fetterman, D. M. (2001). *Foundations of empowerment evaluation.* Thousand Oaks, CA: Sage.

Kirkpatrick, D. (1994). *Evaluating training programs: The four levels.* San Francisco: Berrett-Koehler.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text.* Thousand Oaks, CA: Sage.

Preskill, H., & Torres, R. T. (1999). *Evaluative inquiry for learning in organizations.* Thousand Oaks, CA: Sage.

Smith, N. L. (1994). Clarifying and expanding the application of program theory-driven evaluations. *Evaluation Practice, 15*(1), 83-87.

Evaluating a Sales Training Program

Handout for Activity 21

You are an external consultant bidding on a request for proposals. In your conversation with the vice president of sales, you learn the following:

> The Velocom Company is a global enterprise providing electronic products and services to businesses throughout the world. Because of increasing competition and recent turnover among the sales staff, the vice president of sales and the director of marketing have designed and implemented a new sales training program for all recently hired sales personnel. The sales training is 3 days long and is scheduled to occur twice a year. The program was designed by the two internal trainers who also serve as the program's instructors. The content of the training covers basic sales topics, such as how to begin the sales discussion, how to ask the right questions, and how to ask for the sale. Fifty salespeople have completed the training program in the last 12 months.
>
> Although the vice president of sales and the director of marketing believe this training program is successful, they wish to conduct a more systematic evaluation of its impact on the new sales force. They are therefore looking for an evaluator (evaluation team) to conduct the sales training evaluation.

In Your Groups

1. Discuss each of the evaluation approaches on the handout "Program Evaluation Approaches."

2. Determine which of these approaches you think would best serve this evaluation (you may recommend a hybrid of approaches).

3. Develop a 3-minute presentation outlining your reasons for selecting this particular approach—be prepared to present your idea to the vice president of sales and the director of marketing. (You may use whatever tools or props might be helpful.)

Evaluating a Great
Books Reading Program

Handout for Activity 21

The mission of the Tecolote Group at St. John's College in Santa Fe, New Mexico, is to "acknowledge the dignity of New Mexico's teachers at all levels by providing occasions for building ideas through structured discussion of centrally important texts, thus bringing to life the idea that great books make great teachers." Several different foundations and individual contributors have provided funding for Tecolote.

The Tecolote Group invites approximately 40 teachers per session to attend a series of four all-day Saturday seminars, where they discuss books selected on a particular topic. For example, the theme for the 2003–2004 colloquium series, Phenomenon of Learning, began with Plato's *Meno* and was followed with the work of Aristotle, John Dewey, and Stephen Pinker.

The colloquia are characterized by the following:

- They are designed to be noncompetitive so that participants may explore and learn through shared inquiry.
- They target no particular discipline or grade level but focus on books and questions that all teachers can profitably discuss.
- Participants are chosen with attention to diversity of experience, location, ethnicity, gender, and age.
- All books, other materials, and meals are provided at no charge.
- There is no tuition, and participants receive a modest honorarium for attendance.

The expected outcomes of the program are that teachers return to their classrooms with a renewed sense of their vocation and motivation to develop new initiatives, which they are willing to share with others.

The Tecolote Group has been providing these colloquia for 3 years and has now decided to evaluate the program. They have hired you to conduct the evaluation. The first thing they want to know is what evaluation approach you will use.

SOURCE: February, 2003, colloquia synopsis of The Tecolote Group at St. John's College, Santa Fe, New Mexico.

In Your Groups

1. Discuss each of the evaluation approaches on the handout "Program Evaluation Approaches."

2. Determine which of these approaches you think would best serve this evaluation (you may recommend a hybrid of approaches).

3. Develop a 3-minute presentation outlining your reasons for selecting this particular approach—be prepared to present your idea to the executive director of the Tecolote Group. (You may use whatever tools or props might be helpful.)

Activity 22

Applying Kirkpatrick's Four-Level Approach to Evaluating Training

Overview

This activity helps participants understand appropriate uses of the Kirkpatrick (1994) four-level approach to evaluating training and development programs.

Instructional Objectives

Participants will

- Apply Kirkpatrick's four-level approach to an evaluation case scenario
- Consider and discuss the advantages and disadvantages of using this four-level approach to evaluating training and development programs
- Discuss the reasons why the Kirkpatrick approach has been so popular in the training and development field

Number of Participants

- Minimum number of participants: 4
- Maximum number of participants: 24

Time Estimate: 45 to 60 minutes

In addition to providing the necessary background information on Kirkpatrick's four-level approach to evaluating training and development programs, this activity requires approximately 45–60 minutes, depending on the number of participants and the time available for discussion.

Materials Needed

- Pens/pencils
- Flipchart, markers, tape

- Handout "Sales Training at the Verkauf Company"
- Handout "Variables That Affect the Outcomes of Learning, Performance, and Change Initiatives"

Instructional Method

Case scenario

Procedures

Facilitator's tasks:

- Divide participants into four groups and assign each group one of Kirkpatrick's four-levels: 1. Reactions, 2. Learning, 3. Behavior, or 4. Results.
- Distribute the handouts "Sales Training at the Verkauf Company," "Applying the Kirkpatrick Four-Level Approach to Evaluating Training," and "Variables That Affect the Outcomes of Learning, Performance, and Change Initiatives."
- Ask groups to complete the "Applying the Kirkpatrick Four-Level Approach to Evaluating Training" handout and to write their ideas on a piece of flipchart paper.
- Ask groups to present the information from their flipchart pages.
- Debrief the activity with the following questions:
  - What issues arose as you tried to focus the evaluation?
  - How did you decide which variables your evaluation would focus on?
  - What are the limitations of the Kirkpatrick approach?
  - What is missing from this evaluation approach?
  - Why do you think the Kirkpatrick approach has been so heavily relied on in the training and development field?
  - What could be done to make this approach be more useful, relevant, and effective?

Reference

Kirkpatrick, D. (1994). *Evaluating training programs: The four levels.* San Francisco: Berrett-Koehler.

Sales Training at
the Verkauf Company

Handout for Activity 22

The Verkauf Company is a newly formed organization that provides business services to European enterprises. The vice presidents of sales and marketing decide to offer sales training to all recently hired sales personnel. The sales training takes place as part of the already scheduled regional sales meetings, and the Human Resource Development Department provides the needed trainers. The training covers basic topics, such as how to begin the sales discussion, how to ask the right questions, and how to ask for the sale. Although the two vice presidents believe that this training will be successful, they have requested that you evaluate this training program. You decide to use the Kirkpatrick four-level approach.

Applying the Kirkpatrick Four-Level Approach to Evaluating Training

Kirkpatrick level your group is working on: _____

What aspects of the training will you evaluate?	*What variables will you focus on?*	*What are some potential limitations of the evaluation and its findings?*

Variables That Affect
the Outcomes of Learning,
Performance, and Change Initiatives

Handout for Activity 22

Organizations

- Orientation to change
- Commitment to training and learning
- Resources for training and learning
- Financial situation
- Organizational culture

Trainees

- Motivation to learn
- Readiness for training
- Motivation to transfer learning
- Attitude about and commitment to the job
- Opportunity to apply learning

Trainers

- Facilitation skills
- Content knowledge
- Level of interest/enthusiasm
- Credibility
- Listening skills

Managers

- Ability and willingness to coach on new skills
- Ability and willingness to model new skills
- Expectations of improved job performance
- Provision for time and resources for trainees to use new knowledge and skills
- Communication of the value of training and learning

SOURCE: Adapted from Russ-Eft and Preskill (2001).

134

Training Program Design

- Based on a needs assessment
- Clearly identified training population
- Goals and objectives that are related to identified needs
- Use of a variety of learning/teaching strategies
- Based on adult learning theories and principles

Training Program Implementation

- Appropriate materials to facilitate learning
- Adequate facilities for delivering the program effectively
- Availability of needed equipment
- Adequate schedule for content being delivered

Reference

Russ-Eft, D., & Preskill, H. (2001). *Evaluation in organizations: A systematic approach to learning, performance, and change.* Boston: Perseus.

Activity 23

Debating the Usefulness of the Kirkpatrick Four-Level Approach to Evaluating Training and Development Programs

Overview

This activity engages participants in a debate concerning the strengths and weaknesses of the Kirkpatrick four-level approach to evaluating training and development programs.

Instructional Objectives

Participants will

- Increase their understanding of the Kirkpatrick four-level approach to evaluating training and development programs
- Identify the strengths and weaknesses of the Kirkpatrick approach to evaluating training and development programs
- Discuss how the training and development field can increase its use of other evaluation models and approaches

Number of Participants

- Minimum number of participants: 8
- Maximum number of participants: 24

Time Estimate: 45 to 60 minutes

In addition to providing the necessary background information on the Kirkpatrick four-level evaluation approach, this activity requires 45 to 60 minutes, depending on the number of participants and the time available for discussion.

Materials Needed

- Pens/pencils
- Handout "The Flawed Four-Level Evaluation Model"
- Handout "Invited Reaction: Reaction to Holton Article"
- Handout "Final Word: Response to Reaction to Holton Article"
- Handout "Debating Points"

## Instructional Method

Debate

Procedures

Prior to Class

Facilitator's tasks:

- Make copies of handouts for each participant:
  – Handout "The Flawed Four-Level Evaluation Model"
  – Handout "Invited Reaction: Reaction to Holton Article"
  – Handout "Final Word: Response to Reaction to Holton Article"
- Distribute articles to participants and ask them to read them before the next meeting.

During Class

Facilitator's tasks:

- Assign participants to one of two groups: 1. Pro Four-Level Approach or 2. Con Four-Level Approach.
- Ask participants to take out the Holton and Kirkpatrick articles.
- Distribute the handout "Debating Points."
- Instruct groups to first complete the handout, after which time the debate will begin.
- Explain the rules of the debate:
  – Groups will select four members of their team to be the debaters. The rest of the team will be in a support position. Their role is to provide the debaters with information as needed. This information can be shared only by writing notes—they are not allowed to talk to the debaters during the debate.
  – Each side will have 3 minutes to make an opening statement.
  – For the next 16 minutes, the two sides will take turns making their case. They will each have eight 1-minute turns for making their case or rebuttal.
  – Teams will then have 3 minutes to make their closing arguments.
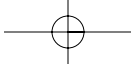
- Debrief the activity with the following questions:
  - What other arguments might have been offered to support your group's position?
  - What other arguments might have been offered to support the opposing group's position?
  - To what extent do you personally agree with Holton or Kirkpatrick?
  - How might the field of training and development reconcile these two positions?
  - What can human resource development professionals do to introduce other evaluation approaches to the field?
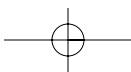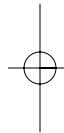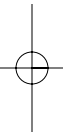
"The Flawed Four-Level Evaluation Model"

Handout for Activity 23

Elwood F. Holton III

The lack of research to develop further a theory of evaluation is a glaring shortcoming for human resource development (HRD). In this paper, I argue that the four-level system of training evaluation is really a taxonomy of outcomes and is flawed as an evaluation model. Research is needed to develop a fully specified and researchable evaluation model. Such a model needs to specify outcomes correctly, account for the effects of intervening variables that affect outcomes, and indicate causal relationships. I propose a new model based on existing research that accounts for the impact of the primary intervening variables, such as motivation to learn, trainability, job attitudes, personal characteristics, and transfer of training conditions. A new role for participant reactions is specified. Key studies supporting the model are reviewed and a research agenda proposed.

Evaluation of interventions is among the most critical issues faced by the field of human resource development (HRD) today. Increasing global competition has led to intense pressure on HRD to demonstrate that programs contribute directly to the organization's "bottom-line." Yet the dominant evaluation model, the four-level Kirkpatrick model, has received alarmingly little research and is seldom fully implemented in organizations (Kimmerling, 1993), leaving them ill-equipped to respond to this pressure. There is a critical need for new evaluation theory and research to give organizations a more sound methodology for allocating HRD resources.

The Kirkpatrick model for training evaluation (Kirkpatrick, 1976), also known as the four-level evaluation model, is acknowledged by many practitioners as the standard in the field. A number of modifications to the model have been suggested, including adding a fifth level to reflect training's ultimate value in terms of organization success criteria, such as economic benefits or human good (Hamblin, 1974) and societal value (Kaufman & Keller, 1994), or to focus more specifically on return on investment (ROI)

(Phillips, 1995). Brinkerhoff (1987) proposed a six-level model that, in essence, added two formative evaluation states as precursors to Kirkpatrick's four levels. Although this work has contributed greatly to our conceptual thinking about evaluation, the models have received incomplete implementation and little empirical testing.

All of them are labeled as *taxonomies,* which are simply classification schemes (Bobko & Russell, 1991). Bobko and Russell, citing Wallace (1983), noted that exploratory designs and case studies are the first steps in theory development, whereas the final steps are correlational and experimental studies. According to them, taxonomies are the link between the initial stages and the final confirmatory stages of developing theory. Although the Kirkpatrick model is elegant in its simplicity and has contributed greatly to HRD, the lack of research to develop further a theory of evaluation is a glaring shortcoming for the field. If HRD is to continue to grow as a profession, an evaluation model grounded in research is necessary.

One shortcoming of taxonomies is that they do not fully identify all constructs underlying the phenomena of interest, thus making validation impossible. Not surprisingly, Alliger and Janak (1989), in their comprehensive review of research on the four-level model, note that the implied causal relationships between each level of this taxonomy have not been demonstrated by research. Their search of the relevant academic literature located only 12 articles since 1959 reporting 26 correlations between levels in training programs out of 203 articles that reported *any type* of evaluation results. Furthermore, only three studies (Clement, 1982; Noe & Schmitt, 1986; Wexley & Baldwin, 1986) reported full four-level evaluations with correlations. The reported correlations varied widely, casting doubt on assumptions of linear causal relationships.

It can be argued that the correlations reported in these studies were not really a test of the model but rather an alternate approach to analyzing outcomes. For example, if only the four levels of outcomes are measured and a weak correlation is reported between levels two and three, all we really know is that learning from training was not associated with behavior change. In the absence of a fully specified model, we don't know if the correlation is weak because some aspect of the training effort was not effective or because the underlying evaluation model is not valid. Weak correlations might represent a well-functioning model reporting a poorly functioning training effect.

It is not surprising that the reported correlations were weak because the model is really only a taxonomy of training (and HRD) outcomes. Attempts to test causal assumptions within a taxonomy are futile because, by definition, taxonomies classify rather than define causal constructions. Kirkpatrick (1994) is unclear about causal linkages in his model. On the one hand, he discusses the influence of other factors such as organizational climate and motivation to learn on training outcomes, suggesting that the relationships between levels are not simple, linear ones. On the other hand, he makes statements that clearly imply a simple causal relationship between levels. For example, he says that "if training is going to be effective, it is important that trainees react favorably" (p. 27) and that "without learning, no change in behavior will occur" (p. 51). The problem is not that it is a taxonomy but rather that it makes or implies causal statements leading to practical decisions that are outside

the bounds of taxonomies. Causal conclusions, which are a necessary part of evaluation, require a more complex model.

Klimoski (1991, pp. 254–256), building upon Dubin (1976), noted that theories or models should have at least six components:

1. *Elements* or *units*—represented as constructs—are the subject matter.

2. There are *relationships* between the constructs.

3. There are *boundaries* or *limits* of generalization.

4. *System states* and *changes* are described.

5. *Deductions* about the theory in operation are expressed as propositions or hypotheses.

6. *Predictions* are made about units.

The four-level model does not meet any of these criteria. First, essential elements are not present. Noticeably absent are the major intervening variables that affect learning and transfer processes such as trainee readiness and motivation, training design, and reinforcement of training on the job (Clement, 1982). Others have proposed models of how individual differences affect training outcomes (Noe, 1986; Noe & Schmitt, 1986) and how factors affect the transfer of training (Baldwin & Ford, 1988; Broad & Newstrom, 1992). Previous evaluation studies identified by Alliger and Janak (1989) did not attempt to measure any intervening variables, which is one likely reason for the wide variation in the correlations reported. No evaluation model can be validated without measuring and accounting for the effects of intervening variables.

Because all of the elements are not present, the relationships between constructs are not fully specified. Considering the third criteria, the four-level model seems to have no limits of generalization within HRD specified. Without full specification of the elements and the relationships, it is questionable whether the model can be applied universally. Furthermore, the missing elements and relationships prohibit making accurate statements about system states, developing propositions and hypotheses, and making predictions.

References

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology, 42,* 331-340.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41,* 63-100.

Bobko, P., & Russell, C. (1991). A review of the role of taxonomies in human resources management. *Human Resource Management Review, 4,* 293-316.

Brinkerhoff, R. O. (1987). *Achieving results from training.* San Francisco: Jossey-Bass.

Broad, M. L., & Newstrom, J. W. (1992). *Transfer of training.* Reading, MA: Addison-Wesley.

Clement, R. W. (1982). Testing the hierarchy theory of training evaluation: An expanded role for trainee reactions. *Public Personnel Management Journal, 11,* 176-184.

Dubin, R. (1976). Theory building in applied areas. In M. D. Dunnette (Ed.), *Handbook of industrial/organizational psychology.* New York: Rand McNally.

Hamblin, A. C. (1974). *Evaluation and control of training.* New York: McGraw-Hill.

Kaufman, R., & Keller, J. M. (1994). Levels of evaluation: Beyond Kirkpatrick. *Human Resource Development Quarterly, 5,* 371-380.

Kimmerling, G. (1993, September). Gathering the best practices. *Training and Development,* 29-36.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook.* New York: McGraw-Hill.

Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels.* San Francisco: Berrett-Koehler.

Klimoski, R. (1991). Theory presentation in human resource management. *Human Resource Management Review, 4,* 253-271.

Noe, R. A. (1986). Trainee attributes and attitudes: Neglected influences on training effectiveness. *Academy of Management Review, 11,* 736-749.

Noe, R. A., & Schmitt, N. (1986). The influence of trainee attitudes on training effectiveness: Test of a model. *Personnel Psychology, 39,* 497-523.

Phillips, J. J. (1995). Return on investment—Beyond the four levels. In E. F. Holton III (Ed.), *Academy of Human Resource Development 1995 Conference Proceedings.*

Wallace, M. (1983). Methodology, research practice, and progress in personnel and industrial relations. *Academy of Management Review, 8,* 6-13.

Wexley, K. N., & Baldwin, T. T. (1986). Posttraining strategies for facilitating positive transfer: An empirical exploration. *Academy of Management Journal, 29,* 503-520.

"Invited Reaction:
Reaction to Holton Article"

Handout for Activity 23

Donald L. Kirkpatrick

I didn't know whether to laugh or cry when I read the title to Holton's article, "The Flawed Four-Level Evaluation Model." After I read the article, I still didn't know how to feel. But when I thought about how Holton had "proved"—through various research papers—that the Kirkpatrick model wasn't really a model at all but only a "taxonomy," I decided not to do either. I will admit that I was a little upset when he listed the six components of a model described by Klimoski (1991) and Dubin (1996) and then stated that "the four-level model does not meet any of these criteria." He might at least have said that it met the second criterion of "relationships between the units" because my model (or my taxonomy, if you prefer) does show the relationships among the four levels.

I admit that Holton's article is a scholarly work. Certainly, cites to other articles are plentiful! Many of them have nothing to do with evaluation, but the list is impressive!

The funny thing is that I personally have never called my framework "a model." Someone else described it that way. For example, in a case study presented in *Evaluating Training Programs: The Four Levels* (Kirkpatrick, 1994), Dave Basarab, head of the evaluation department at Motorola, stated, "Motorola University has adopted the Kirkpatrick model for training evaluation" throughout the world. In another case study presented in the book, the authors from CIGNA Corporation called my model "The CIGNA CMD&T Impact Model." The case study by Patrick O'Hara of First Union National Bank referred to it as "the Four-Level Kirkpatrick Evaluation Model." And Eric Freitag of Intel Corporation wrote an article for the book entitled "Implementing the Kirkpatrick Model as an Up-Front Analysis and Evaluation Tool."

My purpose in offering these illustrations is to demonstrate that, in the real world where training evaluation takes place, the word "model" is commonly used to describe a systematic way of doing something. It may or may not meet the six criteria listed by Klimoski. Personally I don't care whether my work is called a model or a taxonomy as long as it helps to clarify the meaning of evaluation in simple terms and offers guidelines and suggestions on how to accomplish an evaluation. On second thought, I am glad it is not referred to as a taxonomy because if it were, trainers wouldn't know what it meant. The word "model" seems to communicate that it is something to be used as a helpful guide.

I realize that the *Human Resource Development Quarterly* is a scholarly journal that does not publish simple, practical articles. As a former full professor with a Ph.D., I also realize that publishing scholarly articles is required for promotion and recognition in academic circles. I realize too that those who use my model will probably continue to read the *Journal of the American Society of Training and Development,* where the emphasis is on helping the audience rather than on demonstrating the amount of scholarly research done.

Yet I admit I was a little upset by the word "flawed" used in Holton's title. My four-level model has been quoted and used all over the world. Training professionals find it helpful in doing something about evaluation. I have conducted sessions on it at professional conferences in the United States, Venezuela, Mexico, and Singapore. Most recently, in November 1993, I conducted four days of programming for the Arabian American Oil Company in Saudi Arabia.

Indeed, the Holton article tried to tear down the four-level model *without giving any credit to the concepts it depicts.* For example, Holton used the term *reaction,* but not as a level. And the "model" he described used the term *learning,* which is the second level of the Kirkpatrick model. It is interesting that it did not use the term *behavior* but instead a very similar term, *performance.* And, finally, it avoided the word *results* as the fourth level by substituting the words *organizational results.* It seems to me that if my model was so flawed, Holton would not have included so much of it in his concept of a true model.

Finally, in his conclusions, Holton stated that "the model presented here [in his article] is an initial step in the development and testing of a true model of HRD evaluation." He ended by stating that, "If HRD is to grow as a discipline and as a profession, it is imperative that researchers work deliberately to develop a more integrative and testable model." I note that he works at a vocational school. I don't know what qualifications or experience he has with the HRD profession. In contrast, I am a past president of the American Society for Training and Development (ASTD) as well as an author, speaker, and consultant in the field. I regularly conduct workshops at the National Convention of ASTD on the subject of evaluation. These sessions always seem to draw a full house of HRD professionals who are looking for help in evaluating training programs. I only hope that my model—or taxonomy—continues to be of help to the HRD professionals who are more interested in practical ideas than in scholarly research.

References

Dubin, R. (1976). Theory building in applied areas. In M. D. Dunnette (Ed.), *Handbook of industrial/ organizational psychology.* New York: Rand McNally.

Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels.* San Francisco: Berrett-Koehler.

Klimoski, R. (1991). Theory presentation in human resource management. *Human Resource Management Review, 4,* 253-271.

"Final Word: Response to Reaction to Holton Article"

Handout for Activity 23

Elwood F. Holton III

I appreciate Kirkpatrick's reaction to my article "The Flawed Four-Level Evaluation Model." In the spirit of advancing the profession, I will respond to the two issues he raised that relate to practice and research in the profession: the distinction between a model and a taxonomy and the value of evaluation research.

Kirkpatrick states that he never called the four-level framework "a model" and that he doesn't care whether it is called a model or a taxonomy, suggesting that the distinction is merely one of semantics. But the distinction between a taxonomy and a model is critical for both practice and research. If the four levels are truly proposed as a model, then there should be relationships among the levels. Kirkpatrick's response to my article clarifies that he does indeed believe there are. This is helpful because, as I discussed in my article's introduction, his writing has been unclear in this area. If the relationships he refers to are the linear relationships among the four levels he has suggested (see Kirkpatrick, 1994, p. 27 and p. 51, for example) then, unfortunately, the research does not support them. There are variables and relationships missing from the four-level framework, resulting in an underspecified model (Campbell, 1990). If simple linear relationships among the levels are not intended by Kirkpatrick, then a clear specification of intervening variables and relationships is needed. Finally, for the record, he has labeled it a model: "The reason I developed this four-level model . . ." (Kirkpatrick, 1994, p. xiii).

I do not intend to demean the value of taxonomies, which are quite appropriate for intermediate stages of theory development (Bobko & Russell, 1991). They are very useful, and I have developed them myself (Holton, in press). However, it is important to realize that they are just an intermediate stage in making complex phenomena more understandable. Therefore, whether considered a model or a taxonomy, the four-level evaluation framework needs updating. If it is a model, it is underspecified; if it is a taxonomy, then a true model is needed.

NOTE: Elwood F. Holton III is assistant professor of human resource development, Louisiana State University, Baton Rouge.

This issue has critical implications for practitioners, even though they incorrectly use the terms synonymously. The purpose of evaluation is to make decisions about human resource development (HRD) interventions effectiveness and to decide on a course of action to change an intervention if it is not effective. Taxonomies, by definition, do not provide information about causes. Fully developed models do provide causal information, but underdeveloped models lead to wrong decisions or confusion.

Suppose a practitioner finds that learning outcomes in a training program are acceptable but the learning is not used on the job. Then what? Should the training program be canceled? Or should a meeting be called with the supervisors to have them reinforce the training? Should the training be redesigned to include more practical examples? Or should goal setting be included to increase trainee motivation to transfer? The practitioners I work with need to make these kinds of decisions and do not have the tools to do so, even if they implement the four levels.

This issue leads to the second issue I wish to address: the value of evaluation and HRD research. Kirkpatrick suggests in his response that the four levels must be correct because they are widely used. History has shown that the extent to which something is practiced is no indication that it cannot be improved upon. He further seems to suggest that scholarly research articles (such as those published in *Human Resource Development Quarterly*) are not practical or helpful. However, to quote Kirkpatrick, "Progress in evaluation of training will result if all of us will freely exchange information on objectives, methods, and criteria" (1960, p. 17). Isn't that what research does? Is he now suggesting that we should not capitalize on the excellent research that has been conducted to improve upon the four levels by building a more fully specified model? Is he suggesting that we should not conduct research on the four levels because they have been widely used for thirty-five years, essentially without validation or modification? Is he suggesting that the work of scholars in HRD is not important if that work is not first published in *Training and Development?* I hope he doesn't mean to suggest any of these things. Certainly some articles in *HRDQ* may need further refinement before widespread application, but "there is nothing so practical as good research" (Passmore, 1984, p. 24). I have confidence that practitioners in our field are capable of understanding and using models that depict the complex world of human performance in which we work, particularly if they lead to more effective HRD interventions.

In conclusion, if I appear not to have given full credit to Kirkpatrick for the contributions the four levels have made in the thirty-five years since they were first published, then let me acknowledge them right now. As Newstrom (1995) noted, Kirkpatrick is the person who long ago focused practitioners on evaluating outcomes. His four-level framework will always be a classic in the field. But, as Newstrom further noted, HRD research has advanced to the point that we can improve upon the four levels. Furthermore, the complexity of performance improvement practice today demands that we do so. I hope that we can continue to debate and discuss new approaches as scholars in our shared pursuit to improve HRD practice continuously.

05-Preskill.qxd 7/22/2004 5:44 PM Page 148

References

Bobko, P., & Russell, C. (1991). A review of the role of taxonomies in human resources management. *Human Resource Management Review, 4,* 293-316.

Campbell, J. P. (1990). The role of theory in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., pp. 39-73). Palo Alto, CA: Consulting Psychologist Press.

Holton, E. F., III. (in press). New employee development: Review and reconceptualization. *Human Resource Development Quarterly.*

Kirkpatrick, D. L. (1960, Feb.). Techniques for evaluating training programs, part 4: Results. *Journal of the ASTD,* 14-17.

Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels.* San Francisco: Berrett-Koehler.

Newstrom, J. W. (1995). Review of "Evaluating training programs: The four levels by D. L. Kirkpatrick." *Human Resource Development Quarterly, 6,* 317-319.

Passmore, D. L. (1984). Research and theory: There is nothing so practical as good research. *Performance and Instruction Journal, 22*(10), 24-26.

Debating Points

Handout for Activity 23

Group's assignment (either Pro or Con Kirkpatrick's Four-Level Approach): _____

What are the specific points made by the author?	*What does the author recommend?*	*Why does the author make this recommendation?*

Activity 24

Paradigms Exposed!

Overview

This activity engages participants in reflecting on and discussing their worldview as it pertains to inquiry.

Instructional Objectives

Participants will

- Complete a questionnaire that seeks to identify the extent to which they have a more positivist or naturalistic orientation to inquiry
- Discuss how one's worldview may influence an evaluation's key questions, design, and data collection methods
- Consider how their worldview may have affected previous evaluations with which they have been involved

Number of Participants

- Minimum number of participants: 3
- Maximum number of participants: unlimited

Time Estimate: 30 to 60 minutes

This activity requires approximately 30 to 60 minutes, depending on the number of participants and the time available for discussion.

Materials Needed

- Pens/pencils
- Handout "Evaluation Questionnaire"

Instructional Method

Questionnaire

Procedures

Facilitator's tasks:

- Distribute handout "Evaluation Questionnaire."
- Instruct participants to complete the questionnaire to the best of their ability. Explain that it might be difficult to choose one of the two statements in each pair, but not to worry; they should just pick one.
- Emphasize that there are no right or wrong answers and that you will help them score the instrument when they have finished.
- Provide instructions on how to score the instrument: Participants are to give themselves 1 point for each checkmark they have by the following responses (explain that these points do not indicate good or bad or right or wrong):
  - 1a
  - 2a
  - 3b
  - 4b
  - 5a
  - 6a
  - 7a
  - 8b
  - 9a
  - 10a
  - 11b
  - 12b
  - 13a
  - 14b
  - 15a
- Ask articipants to add up their points. Those with a higher number of points lean toward having a positivist worldview, whereas those with fewer points have a more naturalistic or qualitative worldview. Invite participants to raise their hands if they have scores of 0 to 5, 6 to 10, and 11 to 15, respectfully, to see the distribution among participants.
- Debrief the activity with the following questions:
  - What was your experience in completing the questionnaire? Was it difficult? If yes, why?
  - Did your score surprise you? If yes, why? Why not?
  - How might one's worldview affect his or her evaluation practice? How might it have affected your practice with previous evaluations?
  - How would you deal with a client who had a worldview different from yours?

Evaluation Questionnaire

Handout for Activity 24

Directions:

For each of the following 15 paired statements, put a check by the one that comes closest to what you believe. *There are no right or wrong answers.*

_____ 1a. There exists a single reality independent of any person.

_____ 1b. There exist multiple realities that are constructed by people.

_____ 2a. Reality(ies) is governed by immutable natural laws.

_____ 2b. Reality(ies) is *not* governed by natural laws.

_____ 3a. An observer in an organization becomes part of that which is being observed.

_____ 3b. An observer in an organization can remain detached from what she or he is observing.

_____ 4a. The context of a program is needed to understand what is occurring in it.

_____ 4b. A process can be investigated effectively without concern for the specific context.

_____ 5a. Evaluation should be able to determine the true relationship between two variables or factors.

_____ 5b. Evaluation provides tentative conclusions that are always open to interpretation and modification.

_____ 6a. The truth about any relationship between two variables can be determined by testing it empirically.

_____ 6b. The truth about any relationship between two variables can be determined by judgments of knowledgeable experts without further tests.

_____ 7a. Facts and values are independent.

_____ 7b. Facts have no meaning except in some value context.

_____ 8a. Every action is "caused" by an infinite array of considerations that may never be known.

_____ 8b. Every action or outcome has a primary cause that will be identified at some future time.

_____ 9a. The value of evaluation is to predict and control behavior.

_____ 9b. The value of evaluation is to increase understanding.

_____ 10a. Solutions to organizational problems in one organization should be applicable to other organizations.

_____ 10b. Solutions to organizational problems are unique unto themselves.

_____ 11a. Meaningful organizational change is nonlinear and dependent on the active involvement of those affected by the change.

_____ 11b. Change is a rational linear process that will occur naturally regardless of the specific people involved.

_____ 12a. Change is the normal condition of life.

_____ 12b. Change occurs only when something unusual causes it. Nonchange, or status quo, is the normal state of organizations.

_____ 13a. Systematic collection of objective data about knowledge, behaviors, and physical conditions provides the most meaningful knowledge about learning.

_____ 13b. Obtaining feelings, thoughts, and meanings of actions through interviews provides the most meaningful knowledge about learning.

_____ 14a. The more nearly a study reflects the complexity of learning using "thick description," the more valuable it is.

_____ 14b. Quantitative data analyzed with tests of significance are a necessary part of a valuable evaluation study.

_____ 15a. If a study cannot be replicated, and the results verified, I would not have confidence in the study.

_____ 15b. A 6-month case study of an organization carried out by a team of evaluators would provide valuable information even if it could not be replicated and results could not be verified.

Activity 25

Comparing and Contrasting Different Evaluation Designs

Overview

This activity helps participants understand the features of different evaluation designs and when and how each can be used effectively in an evaluation study.

Instructional Objectives

Participants will

- Understand how to determine an appropriate evaluation design
- Identify appropriate data collection methods for various evaluation designs

Number of Participants

- Minimum number of participants: 3
- Maximum number of participants: unlimited when participants are in groups of 3 to 5

Time Estimate: 45 to 60 minutes

In addition to providing the necessary background information on evaluation designs, this activity requires approximately 45 to 60 minutes, depending on the number of participants (or groups) and the time available for discussion.

Materials Needed

- Pens/pencils
- Flipchart, markers, tape
- Handout "Evaluation Designs"

154

Instructional Method

Small-group work

Procedures

Facilitator's tasks:

- Ask participants to get into groups of three to five people.
- Depending on the number of groups, distribute one or two copies of the handout "Evaluation Designs" that depicts different evaluation designs.
- Request that the groups write their ideas on a piece of flipchart paper.
- Ask the groups to share their ideas with the larger group, presenting one or two examples of when their design would be appropriate and useful and one or two examples of when they would not use the approach.
- Debrief the activity with the following questions:
  - As you considered the appropriate uses for your design, what occurred to you? What did you discuss?
  - How would you decide which design to use for an evaluation?
  - Which of the evaluation designs seems most problematic within an organizational context?
  - How would you compensate for the weaknesses of any of these designs if you wanted to use them in an evaluation?
  - How would you educate your client about these different designs if the client was intent on using one particular design that you thought was not useful or appropriate?

Evaluation Designs

Handout for Activity 25

One-Shot Design

In using this design, the evaluator gathers data *following* an intervention or program.

When would you use this design?	*What data collection methods might you use?*	*What are the limitations of this design?*

Evaluation Designs

Handout for Activity 25

Retrospective Pretest

As with the one-shot design, the evaluator collects data at one time but asks for recall of behavior or conditions *prior to,* as well as *after,* an intervention or program.

When would you use this design?	*What data collection methods might you use?*	*What are the limitations of this design?*

157

Evaluation Designs

Handout for Activity 25

One-Group Pretest-Posttest Design

The evaluator gathers data *prior to* and *following* the intervention or program being evaluated.

When would you use this design?	*What data collection methods might you use?*	*What are the limitations of this design?*

Evaluation Designs

Handout for Activity 25

Time Series Design

The evaluator gathers data *prior to, during,* and *after* the implementation or intervention of a program.

When would you use this design?	What data collection methods might you use?	What are the limitations of this design?

Evaluation Designs

Handout for Activity 25

Pretest-Posttest Control-Group Design

The evaluator gathers data from two separate groups *prior to* and *following* an intervention or program. One group, typically called the experimental, or treatment, group, receives the intervention. The other group, typically called the control group, does not receive the intervention.

When would you use this design?	*What data collection methods might you use?*	*What are the limitations of this design?*

Evaluation Designs

Handout for Activity 25

Post-Only Control-Group Design

The evaluator collects data from two separate groups *following* an intervention or program. One group, typically called the experimental, or treatment, group, receives the intervention or program, whereas the other group, typically called the control group, does not receive the intervention. Data are collected from both of these groups only after the intervention.

When would you use this design?	*What data collection methods might you use?*	*What are the limitations of this design?*

Evaluation Designs

Handout for Activity 25

Case Study Design

The evaluator studies an organization or program by collecting in-depth, qualitative data during a specific period of time. This design helps answer *how* and *why* questions and helps evaluators understand the unique features of a case.

When would you use this design?	*What data collection methods might you use?*	*What are the limitations of this design?*

Activity 26

Identifying Evaluation Designs

Overview

This activity asks participants to identify appropriate evaluation designs for various evaluation scenarios.

Instructional Objectives

Participants will

- Understand various evaluation designs
- Determine which evaluation design is being used in three different scenarios
- Discuss the strengths and weaknesses of the designs used in each of the scenarios

Number of Participants

- Minimum number of participants: 3
- Maximum number of participants: unlimited when participants are in groups of 3 to 5

Time Estimate: 30 to 60 minutes

In addition to providing the necessary background information on commonly used evaluation/research designs, this activity requires approximately 30 to 60 minutes, depending on the number of participants (or groups) and the time available for discussion.

Materials Needed

- Pens/pencils
- Flipchart, markers, tape
- Handout "Which Design Is It?"

Instructional Method

Case scenario

Procedures

Facilitator's tasks:

- Tell participants whether they will be working individually or in small groups of three to five people. If they choose small groups, ask participants to get into groups.
- Distribute the handout "Which Design Is It?"
- Instruct participants that, in their groups, they are to discuss each evaluation scenario and determine which evaluation design the scenario represents.
- Tell participants that they will be asked to present their ideas to the larger group and request that the groups draw or represent in some way their chosen designs on flipchart paper.
- Invite groups to share their findings with the large group.
- Debrief the activity with the following questions:
  - What are some of the strengths and weaknesses in each of the evaluation designs you identified in these scenarios?
  - What other information would have helped you make a more informed decision on your choice of design?
  - Which kinds of data would result from each of these designs? What would you know from these data? What wouldn't you know?

Which Design Is It?

Handout for Activity 26

Directions:

Based on the following descriptions, determine which design is being use for each evaluation. Be prepared to explain why you chose these designs and what other designs could be used to evaluate this program or intervention.

1. An evaluation sought to determine the effects of a seminar titled "Giving Effective Feedback" by looking at the differences between seminar trainees (who were self-selected) and a group of employees in similar positions in a department that did not participate in the training. Each group was given a posttest when the training was complete.

2. An evaluation sought to determine the impact of a 4-day workweek on employees' productivity. Measures of employees' productivity were taken prior to implementing the program. The group's productivity was then measured at 3-month intervals for 1 year. At the end of the year, conclusions were made about the program's impact.

3. An evaluation by internal evaluators was conducted to determine the effects of a nonprofit organization's restructuring effort. The evaluators conducted focus group and individual interviews and observed a sample of employees over a period of 2 months. In addition, the evaluators administered an organizational climate survey to all employees.

4. An evaluation compared the effectiveness of providing customer service training on the Web versus in a classroom. Call center staff were randomly assigned to one or the other group. Observations were made of their phone skills 1 month after training. Results were presented to the vice president with recommendations concerning future training.

5. An evaluation examined the degree of transfer of training from a pilot test of a leadership course. First-line supervisors who volunteered for the pilot test were surveyed before the sessions, immediately after the sessions, and 2 months after training.

6. An evaluation focused on a pilot test of a new-employee orientation program. All new employees in the Chicago office participated in the program, and evaluation surveys were distributed at the end of the program. The survey included questions about the facilitator's skills and the usefulness of the content.

Activity 27

Using Evaluation Questions to Guide an Evaluation's Design and Data Collection Methods

Overview

This activity asks participants to identify appropriate evaluation designs for specific evaluation key questions.

Instructional Methods

Participants will

- Understand how an evaluation's key questions influence an evaluation's design and data collection methods
- Identify appropriate evaluation designs and data collection methods for a series of evaluation key questions
- Discuss the relative strengths and weaknesses of each chosen design

Number of Participants

- Minimum number of participants: 2
- Maximum number of participants: unlimited when participants are in groups of 2 to 3

Time Estimate: 30 to 60 minutes

In addition to providing the necessary background information on several evaluation designs (e.g., preexperimental, quasi-experimental, experimental, and qualitative case studies), as well as a basic understanding of data collection methods, this activity requires approximately 30 to 60 minutes, depending on the number of participants (or pairs or triads) and the time available for discussion.

Materials Needed

- Pens/pencils
- Handout "Using Evaluation Questions to Guide an Evaluation's Design and Data Collection Methods"

Instructional Method

Small-group work

Procedures

Facilitator's tasks:

- Ask participants to get into pairs or triads.
- Distributes handout "Using Evaluation Questions to Guide an Evaluation's Design and Data Collection Methods."
- Instruct pairs or triads to read each evaluation question in the left-hand column and to discuss and note, in the right-hand column, which evaluation design and data collection methods might be best suited for answering this question.
- Go down the list of questions, asking for volunteers to share their choices for each question. Ask participants to identify the strengths and weakness of the designs they have chosen.
- Debrief the activity with the following questions:
  - What kinds of things did you consider in choosing an evaluation design?
  - What issues did this activity raise for you?
  - What do you think would happen if two evaluators chose different designs to conduct the same evaluation?

Using Evaluation Questions to Guide an Evaluation's Design and Data Collection Methods

Handout for Activity 27

Directions:

For each of the evaluation questions, choose an evaluation design that would best address the question. In addition, identify one or two data collection methods that could be used to answer the question within the chosen design's framework.

Evaluation Key Questions	*Evaluation Design/Methods*
What impact is the diabetes prevention program having on high school students?	
How does participants' learning from the distance-learning course on customer service compare to participants' learning from the course when it is delivered face-to-face?	
In what ways have program recipients changed their behaviors?	
What impact has sexual assault training had on community service providers?	
How are the program's activities being implemented?	

What are participants' most successful experiences with the program?	
What impact has the new policy had on the ways in which the program is implemented across the 10 sites?	
How are work-life balance policies being implemented and used in this organization?	
What impact is the No Child Left Behind legislation having on curriculum development in Grades K–8?	

Activity 28

Debating the Usefulness of Return on Investment

Overview

This activity involves participants in a debate concerning the strengths and weaknesses of using return on investment (ROI) to determine a program's success.

Instructional Objectives

Participants will

- Increase their understanding of what it means to conduct an ROI evaluation
- Identify the strengths and weakness of using an ROI approach to evaluation
- Determine appropriate uses for an ROI approach to evaluation

Number of Participants

- Minimum number of participants: 8
- Maximum number of participants: 24

Time Estimate: 45 to 60 minutes

In addition to providing the necessary background information on ROI and related concepts, such as cost-benefit and cost-effectiveness designs, this activity requires approximately 45 to 60 minutes, depending on the time available for discussion.

Materials Needed

- Pens/pencils
- Handout "Debating Points"

Instructional Method

Debate

Procedures

Facilitator's tasks:

- Assign participants to one of two groups: 1. For ROI or 2. Against ROI.
- Distribute the handout "Debating Points."
- Instruct groups that they will first have time to complete the handout, and then they will begin to debate the two positions.
- Explain the rules of the debate:
  – Each group will select four members of its team to be the debaters. The rest of the team will be in a support position. Their role is to provide the debaters with information as needed. This information can be shared only by writing notes—they are not allowed to talk to the debaters during the debate.
  – Each side will have 3 minutes to make an opening statement.
  – For the next 16 minutes, participants will take turns making their case or rebuttal. They will have 1 minute each time (eight turns for each team).
  – Each team will then have 3 minutes to make its closing arguments.
- Debrief the activity with the following questions:
  – What other arguments might have been offered to support your group's position?
  – What other arguments might have been offered to support the opposing group's position?
  – To what extent do you personally agree with the position that ROI is the most useful approach to determining the effectiveness or success of a program?
  – Under what conditions might an ROI approach be useful and appropriate?
  – What other kinds of evaluation approaches and designs might evaluators use to determine a program's impact or success?

Debating Points

Handout for Activity 28

Group's assignment (For or Against ROI): _____

The debate focuses on the following statement:

> Determining if ROI is the best approach for evaluating the effectiveness and success of a program.

What arguments can you make to support your position?	*What evidence supports these arguments?*

Activity 29

Is It Really a Return
on Investment Evaluation?

Overview

This activity asks participants to consider whether a return on investment (ROI) evaluation was accurately designed and reported.

Instructional Objectives

Participants will

- Increase their understanding of what it means to conduct an ROI evaluation
- Compare the differences between an ROI evaluation and an evaluation that calculates a program's costs
- Discuss the challenges in conducting an ROI evaluation

Number of Participants

- Minimum number of participants: 3
- Maximum number of participants: unlimited when participants are in groups of 3 to 5

Time Estimate: 30 to 45 minutes

In addition to providing the necessary background information on ROI and related concepts, such as cost-benefit and cost-effectiveness designs, this activity requires approximately 30 to 45 minutes, depending on the number of participants (or groups) and the time available for discussion.

Materials Needed

- Pens/pencils
- Handout "Is It Really ROI?"

Instructional Method

Case scenario

Procedures

Facilitator's tasks:

- Ask participants to get into groups of three to five people.
- Distribute the handout "Is it Really ROI?"
- Instruct participants to read, in their groups, the case scenario on the handout and respond to the discussion questions.
- Invite participants to share their thoughts about the case and their responses to the discussion questions.
- Debrief the activity with the following questions:
  - This case is based on a real example that was disseminated widely on the Internet. Given this, what are your reactions?
  - How could ROI be calculated for this management-training program? What other data would be needed?
  - What challenges might the evaluator have faced if he or she had conducted a real ROI evaluation?
  - How can we help organizations understand the difference between calculating ROI and calculating a program's costs savings?

Is It Really ROI?

Handout for Activity 29

The Situation

A major U.S.-based company had a policy of training several thousand new managers each year. Managers were typically brought together from around the United States, and in some cases from overseas, to centralized locations in the United States for a 1-week management-training event. Two years ago, however, the company began to look for alternative ways of training these managers for several reasons:

- The sheer number of trainees (approximately 5,000 individuals) required a huge administrative staff just to organize and run the 1-week events.
- The cost of bringing thousands of people from multiple locations became a major budgetary issue.
- The increasing complexity of managers' jobs required more than 5 days of training.
- The company was moving into multiple countries. The additional time required to fly to the United States, and the costs of travel, were becoming prohibitive.

The company decided to use a blended approach with e-learning as the primary delivery format for the management training program. In place of the centralized classroom-based training that had been offered before, managers now entered a 6-month e-learning program, which included collaborating online with colleagues to resolve management issues. The training closed with a 5-day, face-to-face session in participants' local offices.

Calculating the Return on Investment

Some time later, the company hired a consultant to calculate the return on investment (ROI) on the new management-training program. The consultant took into account the physical costs of transitioning to e-learning (hardware, software, Internet servers) as well as other costs, such as information technology support, e-learning designers, and content development.

The final report included impressive numbers that showed the following:

- A significant reduction in program, travel, and time away from work costs
- An increase in the amount of content taught during a 6-month period

- A reduction in the cost of course and module development by providing templates for internal business groups to customize their own content rapidly
- A reduction in the time needed to learn (it was estimated that managers could learn the same material in one quarter the time it took using the classroom approach)
- An increase in managers' satisfaction with the blended approach

The consultant's report concluded that the ROI for the e-learning program was more than 2,200%.

Discussion Questions

1. Did the consultant really calculate the program's ROI? If it wasn't ROI, what was it?
2. Was the company's investment in training really paying off in terms of greater productivity and more effective management of teams? Why or why not?
3. Is the ROI figure reliable evidence that the impact on the organization is greater now that managers are trained using a blended approach (e-learning and a 5-day, face-to-face meeting in local offices), rather than the former intensive classroom approach?

Activity 30

Calculating the
Return on Investment

Overview

This activity highlights appropriate uses of return on investment (ROI) and the challenges of conducting this type of evaluation within an organizational context.

Instructional Objectives

Participants will

- Understand the underlying assumptions of conducting an ROI evaluation
- Practice calculating ROI for a given situation
- Discuss situations in which calculating ROI is most appropriate
- Identify the challenges in conducting ROI evaluations within organizations

Number of Participants

- Minimum number of participants: 3
- Maximum number of participants: unlimited when participants are in groups of 3 to 5

Time Estimate: 45 to 60 minutes

This activity requires approximately 45 to 60 minutes, depending on the number of participants (or groups) and the time available for discussion.

SOURCE: This activity was contributed by Barbra Zuckerman, Evaluation Consultant, Albuquerque, New Mexico.

Materials Needed

- Pens/pencils
- Handout "What Is the Return on Investment for Going to College?"
- Small colored stickers (e.g., dots, smiley faces, stars); 80 stickers per group

Instructional Method

Small-group work

Procedures

Facilitator's tasks:

- Ask participants to get into groups of three to five people.
- Explain that they will be calculating the ROI for going to college (either undergraduate or graduate).
- Distribute 80 stickers to each group.
- Distribute the handout "What Is the Return on Investment for Going to College?"
- Explain that participants are to do the following:
  – Identify all of the benefits of going to college and list these in the first column on the handout (e.g., higher paying job, more self-confidence, more opportunities for advancement in career).
  – Identify all of the costs (both financial and nonfinancial) of going to college (e.g., less time with family, more stress, strained work relations) and write these in the second column.
  – Place one to four stickers next to each benefit and each cost, the number of stickers indicating the relative benefits and costs. Participants may not place more than four stickers next to any one benefit or cost. They do not need to use all 80 stickers. The number of stickers relates to the perceived importance or weight of each benefit or cost.
- Instruct groups to add the number of stickers in each column and to indicate the sum at the bottom of the handout.
- Instruct groups to then calculate the ROI given the formula on the handout.
- Invite groups to share some of the benefits and costs on their handout and their ROI results.
- Debrief the activity with the following questions:
  – What was your experience in trying to develop a list of the benefits and costs?
  – How difficult or easy was it to assign the number of stickers to each benefit and cost?

– Did the ROI results you obtained seem credible? If yes, why? If no, why not?
– As you think about applying ROI within an organization, what challenges might you face?
– Under what circumstances might ROI be particularly appropriate? Under what circumstances might it not be appropriate?
– If a client asked you to do an ROI evaluation and you knew it was not an appropriate approach, what would you say? What would you recommend instead?
– What other questions does this activity raise for you concerning the use of ROI in evaluation?

What Is the Return on Investment for Going to College?

Handout for Activity 30

Focus of your ROI study (check one) _____ Undergraduate _____ Graduate

Benefits	Costs
1.	1.
2.	2.
3.	3.
4.	4.
5.	5.
6.	6.
7.	7.
8.	8.
9.	9.
10.	10.

Total benefits (or number of stickers) = _____

Total costs (or number of stickers) = _____

ROI = ([Benefits – Costs]/Costs) × 100%

ROI for going to college = _____