

How Rigorous Should Your Training Evaluation Be?

By James B. Wells

When faced with how to conduct a proper training evaluation, correctional administrators, training directors and trainers must overcome several obstacles, as well as make informed choices from among a variety of different evaluation designs, options and recommendations. Unfamiliar terms and concepts, such as quasi-experimental design, control groups, probability samples, pretests and post-tests, can further complicate matters.

Today, many correctional organizations are faced with budget cuts due to a declining economy and revenue shortfalls. Often the first item to be eliminated from a shrinking budget is training evaluation, followed next by training. However, evaluation becomes even more important when funds are tight. The following are benefits of conducting a training evaluation:

- It improves the efficiency of training, thus reducing costs;
- It provides information on how to improve future training programs;
- It informs decisions on whether to continue training programs;
- It justifies the existence of a training department by showing how it helps achieve the organization's goals and objectives¹; and
- It protects against potential litigation.

All organizations should be concerned with the cost and effectiveness of their training. And in a growing number of correctional organizations, funding sources and upper management are putting pressure on training departments to evaluate more and to produce tangible results from their evaluations. Moreover, those organizations not currently involved in

“failure to train” or similar litigation should take a lesson from those that are: such litigation is becoming increasingly common and more expensive in the justice and safety field. In these situations, trainers need guidelines for evaluating programs, and they need to use more than simple reaction sheets at the end of their trainings. As mentioned in the June 2007 issue of *Corrections Today*, “The days of ‘knowing what I’m doing’ are gone; it takes evidence, not good intentions.”²

Kirkpatrick’s Model

The most popular training evaluation designs today reflect the work of Donald L. Kirkpatrick, who more than 40 years ago described a model for the evaluation of training that involves four levels: 1) reaction; 2) learning; 3) behavior; and 4) results.³ Briefly, reaction refers to determining how well participants liked a training program — for example, by distributing brief evaluation forms at the close of a training event. This is the extent of most training evaluations. Level two involves measuring participants’ learning of facts, principles, skills, attitudes and techniques through some type of pretest or post-test — multiple choice, open-ended, verbal, demonstration, etc. At the third and fourth levels, training evaluation can become more meaningful but also more complex and expensive. Level three refers to the transfer of learning, such as changes in an individual’s job performance, that can be attributed to the training. At level four, the analysis shifts from impact on individuals to impact on the participant’s organization. Conceptually, some have expanded level four to include measuring return on investment (ROI) and cost vs. benefits, while others consider ROI analysis as a level five.⁴

Although it is a simple, straightforward model that training administrators and professionals find appealing, often trainers have problems utilizing the Kirkpatrick model. In Salvatore Falletta’s 1998 book review of Kirkpatrick’s *Evaluating Training Programs*, which was published in the *American Journal of Evaluation*, he said, “Everyone’s talking about it. No one is doing it.” Why is that? Simply put, a Kirkpatrick-style evaluation is “too much” for some training situations and “not enough” for others, and it provides no guidelines for determining if it’s appropriate, or what to do if it isn’t.

Not all trainings should be evaluated to the extent Kirkpatrick recommends. To illustrate, in describing guidelines for measuring each of the levels, Kirkpatrick recommends pretests and post-tests for measuring learning and, if feasible, that a control group be utilized to assess learning, behavior and results. However, depending on the particulars of a given training, it may be neither desirable nor necessary to perform all the evaluation steps he recommends. For example, an inexpensive half-day training program designed to orient staff to a new internal policy that minimally affects their duties may not deserve as much evaluation rigor as an expensive, week-long training program designed to reduce institutional violence by training staff in improved interpersonal communication skills. Whereas a simple level one reaction sheet may be sufficient to evaluate the half-day internal policy training, a more rigorous evaluation that attempts to measure knowledge and behavioral and organizational change may be needed to adequately evaluate the week-long skills training. In fact, when more thorough evaluations are warranted, a Kirkpatrick-style evaluation may be inadequate.

While the Kirkpatrick model includes some important process measures (reaction and learning) and outcome measures (behavior and results), it fails to include several other important measures. For example, some evaluations will require a needs assessment to determine if there are training needs currently not being met that the training program will satisfy. Likewise, an evaluability assessment will help determine the extent to which it would be productive or feasible to evaluate a program. Another important part of a thorough training evaluation absent from the Kirkpatrick model is the program implementation assessment, which evaluates the extent to which the program was implemented as designed (e.g. sufficient resources, competent staff, etc.). The target audience assessment is used to determine if the training is being delivered to the people for whom it was designed. Finally, a cost-efficiency or ROI assessment provides important information about the total costs of training (in dollars, downtime, etc.) relative to immediate and long-term training results, for both the individual training participant and the participant's organization. These and other elements of a thorough training evaluation are illustrated in Table 1. Note that the Kirkpatrick model includes only four of the nine assessments required for most rigorous evaluations. These nine assessments were developed, using established literature on evaluation research and training evaluation, by a team of three researchers consulting with NIC under the umbrella of Commonwealth Research Consulting Inc.⁵

For those unfamiliar with process and outcome measures, process measures examine the extent to which a program was implemented as designed and is operating as designed. Good process evaluation work is crucial. It can help identify, at early stages of program implementation, things that are going to enhance success and things that are going to be barriers. Outcome measures, on the other hand, examine the impact of the training (i.e., the extent to which the program promoted positive individual behavioral change and organizational change in an efficient and cost-effective manner).

Table 1. Elements of a Thorough Training Evaluation

Preliminary Measures	Process Measures	Outcome Measures
Needs assessment	Target audience assessment	Behavior (Level 3)
Evaluability assessment	Program implementation assessment	Organizational results (Level 4)
	Reaction (Level 1)	Cost-efficiency or ROI
	Learning (Level 2)	

Rigor Tests

The nine evaluation components shown in Table 1 are required for more rigorous evaluations. Rigor refers to the degree of confidence desired for the reliability (i.e., consistency) and validity (i.e., accuracy) of evaluation findings. To the extent that an evaluation involves more advanced rigor, there is greater corroboration of information obtained, more attention to reliable and valid measurement, more complex data analyses, and greater effort to track and control influences that might bias results. In general, the more advanced the rigor, the greater the confidence that can be placed in the results. The degree of rigor is determined by the questions of interest as well as the designs, measures and other methods needed to address those questions.

While numerous levels of rigor are possible, consider for a moment just the basic and advanced. In a basic process evaluation, reaction and learning could be measured with a simple, nonvalidated form that asked participants a series of questions about how much they liked various aspects of the training (without seeking objective measures of training quality) and how much they learned about various topics (without testing them on the learning). Basic outcome evaluations would likely consist of participants' self-ratings of individual behavior change or organizational change using nonvalidated instruments. Cost-efficiency or ROI evaluations would merely include self-ratings by stakeholders and include estimations of cost savings, rather than relying on data collection to assess actual cost savings.

More rigorous, advanced evaluations are warranted in some cases. Keep in mind, however, that as the degree of rigor increases, so does the complexity and cost of the overall evaluation. Given such factors as scope of the training, its intensity,

length and projected cost in both monetary and human resources, as well as commitment of the participating agencies and their personnel, a variety of advanced evaluations may be required. In an advanced evaluation, the breadth of evaluation may need to be expanded to include all nine assessments, including a thorough process evaluation to collect and analyze a large amount of target audience and program implementation data, as well as data on the training participants' reactions to the training and knowledge gained from the training. Additionally, outcome evaluations may be needed to gather data pertaining to whether the training influenced any change in the behavior of the participants or the organizations to which they belong. Last, advanced evaluations will likely also conduct pre and post cost-benefit and/or cost-effectiveness analyses by estimating costs associated with the program and comparing that to data on cost savings pertaining to program benefits.

A New Model

A major byproduct of Commonwealth Research Consulting Inc.'s training evaluation conducted over several years is a modification and expansion of the Kirkpatrick Model. In an ongoing training evaluation project with the National Institute of Corrections (NIC), the team and NIC decided to address the implementation problems of the Kirkpatrick model by developing an improved evaluation model. This model is called the Evaluation Matrix.⁶ The Evaluation Matrix, when employed in conjunction with training objectives, will assist correctional staff in identifying the appropriate type of evaluations to conduct as well as the proper level of evaluation rigor. The Evaluation Matrix improves on the four-level model by:

- Expanding the breadth of the evaluation to include additional key elements such as needs, evaluability, target audience, program implementation and ROI assessments;
- Expanding the depth of the evaluation to include selection of necessary and appropriate levels of evaluation rigor; and
- Providing a tool to help plan an evaluation strategy.

The last point is the most important because what works for one agency may not work for another. The matrix is a work in progress, subject to continual refinement based on feedback and use. An important goal in this respect is to make it more useful to people not familiar with evaluation research. The research team and NIC hope to make it available soon to the field to help leaders in designing and conducting evaluations.

ENDNOTES

¹ Kirkpatrick, D.L. 1996. *Evaluating training programs: The four levels*. San Francisco: Berrett-Koehler Publishers.

² Potter, H.R. 2007. From good intentions to evidence-based: Paving the right road. *Corrections Today*, 69(3):74-75.

³ Kirkpatrick, D.L. 1959. Techniques for evaluation programs. *Journal of the American Society of Training Directors (Training and Development Journal)*, 13(11):3-9. Available at www.emeraldinsight.com.

⁴ Philips, J.J. 1996. ROI: The search for best practices. *Training and Development*, 50(2):42-48.

⁵ Wells, J.B., K.I. Minor, L.H. Wallace and J.S. Parson. 2007. *National Institute of Corrections Training Evaluation Project: Selecting evaluation type and rigor*. Washington, D.C.: U.S. Bureau of Prisons, National Institute of Corrections.

⁶ Ibid.

James B. Wells, Ph.D., is a professor in the Department of Correctional and Juvenile Justice Studies at Eastern Kentucky University and president of Commonwealth Research Consulting Inc. Wells is a former correctional officer at a maximum-security institution.