Evaluating Distance Delivery and E-Learning

IS KIRKPATRICK’S MODEL RELEVANT?

by Dominique L. Galloway

Just as technology has dramatically changed the environment in which businesses operate, it has also changed the way businesses deliver training to employees. Integrating new technology with experiential education can substantially enhance the learning process (Bonk, 2002). Thus, distance delivery, including e-learning, offers corporations a way to improve training while increasing performance initiatives and potentially delivering higher returns on training investments. Assessing the outcomes of training has always been challenging, and distance education presents new complications to evaluating e-learning. Although e-learning offers promising results and reduced costs, traditional methods of evaluating training and learning must be adapted and re-examined with an eye to their applicability to e-learning environments.

One of the most dominant schemas for evaluating the effectiveness of training programs is Donald L. Kirkpatrick’s (1998) four-level model used to evaluate instructor-driven, classroom-oriented corporate training programs. Although it was developed more than 30 years ago, instructional technologists continue to use Kirkpatrick’s method for evaluating more traditional classroom-based training programs. This article addresses important questions that have been raised as to how accurate this method is in assessing technology-based training:

- Is the Kirkpatrick four-level model an optimal evaluation approach for e-learning?
- Are there alternative evaluation models in education, the social sciences, and business that might prove more useful to instructional technologists seeking more accurate ways to evaluate distance education and distance-oriented performance-improvement interventions? (Because they may differ from traditional learning, can they be evaluated similarly?)
- What would an ideal evaluation model for distance situations look like?
- What are the gaps between an ideal model and the current evaluation models?

Kirkpatrick’s Four-Level Model

The effectiveness of any training endeavor can be measured using any outcome that an organization deems relevant; however, evaluation on only one set of criteria can lead to biased conclusions. To minimize such bias, Kirkpatrick (1987) identified four levels for evaluating training: (1) reaction, or how well learners were satisfied with the program; (2) learning, or the principles, facts, and techniques that were learned; (3) behavior, or the changes in job behavior that resulted from the program; and (4) concrete results in terms of cost reduction and quality and quantity improvements.
Although Kirkpatrick's model has been the dominant method of evaluating training programs for more than 30 years, criticisms of his model and its shortcomings nevertheless exist. Aldrich (2002) notes that Kirkpatrick's model might be too simplistic, and he has several reasons for this critique, two of which are relevant here. First, Kirkpatrick assumes that evaluations are definitive and can be replicated and generalized to larger groups. But as Bennett and Arthur (1997) note, many factors influence training effectiveness in organizations, and all these factors simply cannot be accounted for by any one model. Second, Kirkpatrick's evaluation of effectiveness focuses just on whether the training has produced the intended outcomes. However, training effectiveness must be evaluated by considering several organizational, individual, and training-related variables that may affect outcome, including the evaluator's knowledge of and preparation for the training topic (Aldrich, 2002).

Abernathy (1999) contends that despite the shortcomings of Kirkpatrick's model, it does offer some flexibility. This flexibility specifically applies to Kirkpatrick's (1998) encouragement of trainers to borrow approaches, techniques, and methods from others and to understand the difference between proof and evidence of training results. Desired results—such as greater progress in accomplishing critical goals, improved service to customers, increased productivity, and greater job satisfaction—measured through company reports and statistical averages can verify improved processes, clarity of goals and priorities, greater commitment of employees, and more efficient operations due to training. Strategic alignment is the degree to which a training program conforms to and advances the goals of the organization for which it is designed (Hancock, 2004).

This has specific bearing on the Kirkpatrick model because that model is based on a systematic matching of goals and outcomes—all of which can be determined and customized by a company and its trainers before training begins. Although strategic alignment is not an explicit consideration within the Kirkpatrick model, trainers—to enhance strategic alignment, particularly in the second, third, and fourth levels of the Kirkpatrick model—can easily extrapolate from the model. The ease with which strategic alignment can be evaluated within the Kirkpatrick model has contributed to the model's popularity and perseverance as the dominant model of evaluation.

**Level One: Reaction**

In Kirkpatrick's evaluation model, the first evaluative level is participant reaction. To provide tangible measurement results at this level, participants often complete a questionnaire at the end of a training session. This type of reactive evaluation gives training participants the opportunity to evaluate various aspects of the training, including the topic, the quality of the materials used, the manner in which the topic was presented, the length of the course, and the quality of the instructor. In practice, some training professionals rely strictly on this level of evaluation and bypass the other three levels (Morrow et al., 1997). Although over-reliance on any level of evaluation can be problematic, the reaction level nevertheless provides immediate information by helping managers and supervisors determine whether the participants find the program valuable.

Reaction evaluations cannot exclusively be relied on to provide an accurate measure of a training session's effectiveness because they have specific limitations. For example, the first level of Kirkpatrick's model cannot definitively determine whether a program has produced a return on the sponsoring firm's investment. Additionally, it does not objectively demonstrate that the participants in the training session learned or internalized the goals of the session. Davis and colleagues (1998) refer to these types of evaluation forms as "smile sheets." A description that highlights the subjective nature of this evaluation, while Aldrich (2002) dismisses evaluations that rely exclusively on Level One data as irrelevant.

Aldrich further cautions that Level One evaluation responses may cause organizations to fruitlessly revise programs in response to feedback that may have had nothing to do with the actual training. Myriad factors may influence negative feedback from trainees, including lack of interest in the topic, distraction by personal problems, resentment of the time the training takes, or even the fact of required participation. For these reasons, the remaining three levels in the Kirkpatrick model should be considered to generate more holistic and accurate evaluations.

**Level Two: Learning**

Trainers use the second level of the Kirkpatrick model to determine whether the training has led to an increase in knowledge or skills. A number of techniques can be used to assess learning: one such way is to ask participants what their learning expectations were for the training and then to ask if those expectations were met. A substantial limitation to this type of evaluation, however, is again that these responses are subjective and may not even express an employee's true feelings—that is, the response may be based on how the person feels about the training itself, or a positive response may be based on what the employee thinks the trainer wants to hear. Another is to test learners on instructional content. In any event, learning evaluations at Level Two of the Kirkpatrick model still do not indicate whether any meaningful transfer of learning or skills has taken place. For example, a trainee who demonstrated in a training session that he or she understood and could apply concepts of employee counseling in role-playing situations might not be able to do so in a real situation.
Those considerations are known as barriers to transfer and are a constant concern in the execution and evaluation of any training session. According to Ford and Weissbein (1997), only 10%–30% of what is taught in training programs actually translates to employee knowledge and skills. Esque and McCausland (1997) report that its employees retained less than one percent of one Fortune 500 Company's training.

The Kirkpatrick model or any other model used to evaluate participants’ learning and retention must acknowledge barriers to transfer in the final evaluation of a program’s success. Although the Kirkpatrick model provides more substantive evidence in Level Four, Level Two relies heavily on evaluation data provided by participants and trainers. This is potentially problematic, because self-assessment is rarely the best way to determine whether a person’s behavior actually has changed because of training. To be accurate indicators of a training session’s effectiveness, data sources should be unbiased, provide understandable information, be easy to access, and produce information that is immune to irrelevant influences (Wisher et al., 2001). In this way, the first two levels of the Kirkpatrick model are useful but are not often substantiated by hard facts. Therefore, they are subject to bias. Because these two levels are unresponsive to barriers to transfer, they are potentially unreliable indicators of the effectiveness of the training session.

Level Three: Behavior

A vital part of establishing the utility of the training investment is the valuation of training programs for the transfer of knowledge, skills, or aptitudes to the job. The third level of the Kirkpatrick model, which evaluates behavior and how the training session translates into employees’ knowledge and aptitude, seeks to measure the continuity between learning and practice. According to noted evaluation expert Jack Phillips (1996), evaluating at Level Three of the Kirkpatrick model servers the following functions: (1) determining success in accomplishing organizational goals, (2) identifying strengths and weaknesses in the training and development process, (3) identifying which participants were the most successful, and (4) providing an opportunity to reinforce major points to the participants.

Like Kirkpatrick, Phillips stresses the importance of measuring change on the job caused by the application of training material. Often, however, the supervisor is asked to provide such evaluative data, creating potential obstacles to the successful administration of Level Three evaluations. In the flattened organizational structure common to most businesses today, supervisors often do not have the time or the expertise to effectively evaluate change. They may additionally have an emotional link to the people or situations being evaluated, creating biased results. Because a leading criticism of the Kirkpatrick model contends that it has been in place with little modification since 1959 and is thus potentially out of date, the weaknesses or strengths of Level Three are relevant to that discussion. As an evaluative measure, Level Three does not necessarily need to be conducted at a particular location or even in person. Technology allows various attributes and aspects of job performance to be assessed using networked computers. This enables some of the measurements in the Kirkpatrick model that were once done in a physical place to take place online, as long as the computer-based elements of the evaluation do not take place too far in advance or too long after the rest of the assessment. Because computer-based performance tests can measure declarative knowledge (e.g., a list of facts) and procedural knowledge (e.g., steps in a procedure), computer-based performance testing can be used to “demonstrate that a person can apply their knowledge and skills to perform tasks, take actions, and solve problems that are realistic indicators of proficiency or competency” (Berry & Malone, 1999, p. 164). This helps eliminate much of the potential bias noted earlier.

Level Three has the potential of being more relevant to evaluating today’s training (including e-learning) than it did when the model was first implemented and technologies such as computers were less available to measure learning. As Figure 1 illustrates, computer-based performance testing provides a promising and accurate gauge of participants’ learning and understanding, as evidenced by metrics that make up a computer-based portion of the evaluation are shown. In distance learning programs, it is possible to measure whether assignments were completed correctly, whether they translated into learning through tests, and whether they were completed within the time restraints. Therefore, computer-based performance testing like that shown in Figure 1 can be used for evaluating computer-based tasks, such as the use of a particular software program while completing a specific task. This application of Level Three was not available until after widespread use of personal computers. Computer-based testing is the most promising adaptation of the Kirkpatrick model to e-learning, as it provides both hard data—that is, scores, rates of achievement, and aptitude—and tangible results including but not limited to performance comparisons and cost effectiveness for evaluation and compilation for further study including cost effectiveness and is easily integrated into today’s distance-learning environment.

<table>
<thead>
<tr>
<th>Process</th>
<th>Were the correct tasks performed?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence</td>
<td>Were the tasks performed in the correct order?</td>
</tr>
<tr>
<td>Results</td>
<td>Were the correct results obtained?</td>
</tr>
<tr>
<td>Time</td>
<td>Were the results obtained within time constraints?</td>
</tr>
</tbody>
</table>

Figure 1. Sample Metrics to Evaluate E-Learning (Note: These metrics expand on Kirkpatrick's Level Three, and the questions can be changed or customized to meet the evaluator's needs or to act as a component of strategic alignment.)
This particular knowledge is fundamentally tested through computer-based test application with regard to program use. Based on a directive given, the test evaluator can see whether the applicant performed the desired steps, trace the order of the steps, and view the finished product and/or result. Time constraints can also be monitored. Therefore, this method can truly measure the learning attained.

**Level Four: Results**

In the Kirkpatrick model, Level Four is the level most anticipated by management because it produces evidence that can be related to increased sales, reduced costs, increased productivity, improved quality, and lower overhead. This type of hard data can be used to justify or improve additional training and development efforts. Level Four is the most important value level, as it attempts to assess training in business results and in terms that managers and executives can understand, that is, increased production, improved quality, reduced costs, reduced frequency of accidents, increased sales, and even higher profits or return on investment (ROI). From a business and organizational perspective, it is clearly the most difficult data to obtain (Winfrey, 1999). In a traditional classroom or a computer-based classroom, information about the participants’ learning is essential for evaluating both the effectiveness of the training and the cost-benefit of the training to the sponsoring company. E-learning and distance learning’s relative newness often causes those who are more conservative to be skeptical about whether this type of training actually results in any learning or behavioral changes and thus productivity and profit (Martinez, 2003).

While Level Four of this model is highly relevant to distance learning, as it provides tangible evidence that learning has occurred, it validates Level Two. A supplemental model, the ROI method discussed below may act as an additional measurement of effectiveness of a given training session. Although the Kirkpatrick and ROI models are superficially similar, ROI produces results that specifically outline a given training program’s monetary return, whereas the second level in the Kirkpatrick model generally evaluates learning results rather than monetary returns Level Four reports. The newness of e-learning, like any novel business initiative, opens itself to potential skepticism with regard to effectiveness, necessity, increased asset value, cost effectiveness, and increased production. Therefore, it necessitates an evaluation model that provides evidence of learning as well as financial results. In today’s business environment, any type of training—whether it is traditional or technologically based—must demonstrate cost-effectiveness (Martinez, 2003). In this way, a combination of the Kirkpatrick and ROI models is able to uniquely meet the consumer and learning demands of distance education.

**ROI as a Supplemental Evaluation Model**

Whereas the Kirkpatrick method provides for four levels of evaluation, Phillips (1996) adds a fifth level to address the recurring need to measure the ROI for training and development activities. While some training professionals argue that measuring ROI for training is not possible (Bennett & Arthur, 1997; Lupton et al., 1999), there is general agreement (Anderson, 1993; Truelove, 1997) that trainers must show an ROI so that funds will continue to be made available for training programs. The credibility of a training program can be enhanced within the organization if an ROI can be demonstrated. To determine the net result, Phillips’ added a fifth level to Kirkpatrick’s model, which focuses on the assessment of the monetary value results yield, compared to the cost of training. In other words, this creates a cost-benefit model, in which monetary benefits are compared to cost to determine whether training costs were excessive.

As an example of ROI measurement after training, Phillips (1996) describes the ROI in a literacy-skills training program conducted at Magnavox. The program produced benefits of $21,600, and the cost was determined to be $36,233. Thus, the net benefit of the training was $283,367, which represents a 74 1/4% net benefit to the company. In this case, Phillips’ fifth level provides clear evidence that there was a return in net benefits for every dollar Magnavox invested in that training program. Phillips’ fifth level substantially increases the amount of hard evidence available to evaluators and the companies funding training and development. In this way, a modified Kirkpatrick system is able to evaluate the outcomes of training with a full complement of subjective and objective data, making it as relevant to a technology or computer-based classroom as it is to traditional classroom instruction.

In terms of assessing the cost-effectiveness of training and, specifically, the cost-effectiveness of e-learning, Levin may provide significant insight in the statement that “it is important to emphasize that the evaluation of effectiveness is separable from the evaluation of costs” (1995, p. 382). This observation is relevant to a consideration of ROI in an e-learning environment because it allows evaluators to consider a program’s cost and effectiveness separately. Although this separation may be only momentary and does not absolve a program’s accountability for profits, it nevertheless may allow trainers and evaluators using the ROI method to evaluate a program’s merit independent of its cost. Should a component of an otherwise expensive program be worthwhile, evaluators using Levin’s approach may be able to salvage it. This is further applicable to a technology-based classroom as companies and trainers continue to seek ways of adapting the effective parts of conventional teaching to the more often cost-efficient advantages of e-learning.

If calculating ROI for training and development has been a persistent concern, it follows that calculating the relative
ROI of e-learning is equally important (Docent, 2002; Reddv. 2002). In an e-learning environment, unique costs and benefits exist and must be accounted for in an accurate measure of ROI. For example, e-learning training and development typically involves reduced travel expense, reduced time away from the workplace, 24-hour, 7-day-a-week access to training, and possibly reduced turnover. In this way, e-learning’s ROI can be defined in the same terms of dollars, benefits, and outcomes that define traditional learning (Docent, 2002; Horton, 2001).

To develop an accurate and valuable ROI analysis for distance learning, a number of direct and indirect costs of program development, delivery, and maintenance must be considered. Some direct costs include trainers’ compensation; outside vendors; development, production, and distribution of materials; hardware; software; and administrative support. Indirect costs include the duplication and distribution of CDs to trainees and the purchase and maintenance of a computer server for web-based training (Kruse, 2002).

The next step in this expansion of the Kirkpatrick model is determining the final details of direct and indirect costs associated with the administration, maintenance, and technological updating of the training program. The final step involves recording all the subtotals gathered to create a comprehensive picture of the total cost comparison. However, where do Kirkpatrick’s Levels Three and Four fit? What specific relevance do Levels Three and Four have when the ROI model is applied to evaluating distance delivery or e-learning efforts?

There are no easy answers to these questions. The entire context within which the training occurs and how the training fits within the organization’s operations and culture or strategic alignment are important to evaluate, especially in evaluating any measurable changes from Level Three or Level Four. For example, lack of management support can undermine even the most effectively designed and delivered training program. As a potential remedy to this, Davis and colleagues (1998) recommend establishing a link between training interventions and such outcomes as job behaviors and productivity so that measurable changes resulting from training can be identified.

Most evaluations can benefit from measuring organizational context. Marquardt (1996) suggests that evaluation that the Kirkpatrick model provides is deficient because an evaluation should measure more than reaction, learning, behavior, and results, as those attributes focus mostly on outcomes and not on the processes leading to the results. For example, training does not occur in isolation from an employee’s job responsibilities or personal life, so one can evaluate behavior effects without recognizing that they might depend on people’s motivation, the degree of managerial support after training is completed, or the extent to which the training was appropriate for meeting needs. Furthermore, simply providing time for employees to receive training does not ensure training effectiveness (Mathieu et al., 1993). Considerations such as these go beyond the monetary and are not necessarily included in either the ROI or Kirkpatrick models, even though they are important to a complete and accurate evaluation of a training program’s success.

Discussion

There are almost as many ideas about how to measure the effectiveness of training as there are trainers. Whether e-learning takes place successfully still depends on the delivery system used, the motivations and desired behaviors, the end results, and the underlying e-learning use.

Kirkpatrick’s four-level evaluation model is hardly dead, but it is being challenged or at least refined (Holton, 1996), particularly as it relates to e-learning (Allen, 2003). While the model continues to be the most popular framework for categorizing training criteria and provides a simplicity that is quite appealing, it contains a number of assumptions that can lead to overgeneralizations and misunderstandings that compromise the evaluation of e-learning.

The most ideal model for evaluating technology-based learning may simply be an enlightened synthesis of the Kirkpatrick and ROI models. Although Shelton and Alliger (1993) provide a practical approach for completing the Kirkpatrick four-level evaluation challenge and avoiding the overgeneralizations that frequently result, an ideal model would include a method for evaluating intellectual property in terms of employee expertise and skill levels. This would substantially aid an employer in determining the worth of an employee or group of employees beyond merely using employee benefits and compensation rates. A hybrid of the Kirkpatrick and ROI models is best suited to evaluating learning because it would include both data such as an employee’s application of knowledge, which, although less tangible than other forms of data, is still essential, as well as the specific dollar return on a company’s investment in training.

Such a model for evaluating learning must take into consideration ways to assess technological innovation in training and in skills, especially in today’s highly competitive, technologically sophisticated global marketplace. The value of technological innovation as a factor in a company’s growth has changed the way the ownership of intellectual property is regarded, especially in the past decade. Computer programmers, for example, commanded salaries around $80,000 during the early to mid 1990s in the United States because they were innovative and educated. They possessed the intellect the companies needed to forge into the next millennium. In the 21st century, intellect is valued differently because the Information Age and international trade has raised standards and increased the demand for innovative
solutions along with cost-effective production and administration. Because the knowledge gained from the technology and those trained in its use are perceived as valuable corporate assets—that is, intellectual property—companies want to preserve that asset. Therefore, preservation of accrued and improved human assets should prevent the competition from luring one of it most value assets, a competent, trained employee. This represents another gap between ideal models and current ones, which leaves the question of how one can evaluate the acquisition of this type of skills and expertise from a training program.

Training professionals often do not like to acknowledge the historically low stature and importance that training and learning occupy within the corporate hierarchy because of the perceived difficulty of determining ROI for training activities. This is due, in part, to the inability of some training professionals to make a business case for training. E-learning ROI offers the potential to do just that and to change dominant perceptions of the role of learning in the workplace. However, it is important that management perceive the importance of e-learning in the organization and the use of such learning to increase revenue. If such a realization is attained, managers would be more likely to value e-learning’s contribution to their business goals, and e-learning could rapidly become a strategic corporate tool.

Although an ROI approach provides the most promising method of evaluating the outcomes of technology-based learning, the implementation of ROI to evaluating training, however, does not necessarily mean we have to abandon Kirkpatrick’s model. A blend of Kirkpatrick’s four-level model and the ROI model adapts Kirkpatrick’s 30-year-old method to the demands of a more technologically based training environment. Such an adaptation satisfies the need of businesses and managers to evaluate the bottom line in training and development programs while simultaneously providing an ideal method for effectively evaluating distance-delivery and e-learning activities. Computer technology makes the Kirkpatrick model more relevant, not less. At the same time, competitive business practices demand cost-effectiveness and a measurable ROI in training, making a way of determining training’s cost-to-benefits ratio essential. In this way, the Kirkpatrick and ROI models are strengthened through a hybridization of their different approaches, and this hybrid is uniquely responsive and adapted to evaluating e-learning in today’s market.

References


**Dominique L. Galloway** has a BS in Marketing from Ball State University as well as MS degrees in Information and Communication Sciences (Ball State University) and Instructional Systems Technology (Indiana University). He is currently pursuing an Instructional Systems Technology doctorate at Indiana University. He is a Business Research Analyst within Merck. He is responsible for market research activities to support strategic and tactical decisions for Vaccines, Customer Marketing, and Policy. He joined MVI in 2001 as an Instructional Designer as part of the sales training and development team. In this position, Dominique led the effort to evaluate and redesign the basic training program. He also developed a basic training instructional guide for district managers for use with their teams and has improved efficiency in this department by implementing computer-based testing. Prior to joining MVI, Dominique spent 12 years with Eli Lilly and Company where he held positions as a Trainer, Instructional Designer, Sales Representative, and Systems Analyst. He may be reached at dominique_galloway@merck.com.

**IPMA-HR is the leading organization addressing the needs of public sector human resources professionals.**

By offering the tools, resources, and information unique to the public sector, IPMA-HR works to enhance the image of the human resources professional through ongoing professional, personal and organizational development.

**For more information contact IPMA-HR at 703-549-7100, or visit www.ipma-hr.org**