Evaluation and eLearning

Debra PEAK
Zane L. BERGE

UMBC
1000 Hilltop Circle
Baltimore MD 21250

ABSTRACT

In today's results-oriented, fast-moving business environment, it is critical for trainers to demonstrate the value of training to the organization: There is nothing inherently valuable about training. It is performance gains that training catalyzes that give it worth (Graber, 2000). This is why evaluations tied to business results are becoming commonplace.

If you ask training professionals about measuring training, most will start talking about levels of evaluation, referring to Kirkpatrick's landmark evaluation model developed in 1959. Kirkpatrick's levels of evaluation have been the industry standard for nearly half a century.

However, many professionals now believe that elearning and a shift in emphasis toward performance improvement have changed the training business so that these levels are no longer completely relevant.

The purpose of this paper is to discuss what similarities and differences exist between evaluating elearning and traditional classroom instruction, how Kirkpatrick's evaluation levels are currently conducted, why conducting Kirkpatrick's Level 4 evaluation is so difficult to do, why elearning evaluation has evolved to include return-on-investment (ROI) calculations, and whether other evaluation methods currently practiced are more relevant and useful.

Keywords: **eLearning, Evaluation(s), return-on-investment (ROI), traditional classroom instruction, Kirkpatrick's levels of evaluation.**

EVALUATING eLEARNING VERSUS IN-PERSON CLASSROOM INSTRUCTION

In some ways, elearning is simply another method of delivering instruction. So, it should be no surprise that there are many similarities in the evaluation process. Both traditional and elearning instruction collect the same types of data (hard and soft, quantitative and qualitative); use at least some of Kirkpatrick's four levels and occasionally Phillips' fifth level; use some of the same data collection methods needed to convert soft data to monetary form; and analyze and report the data the same way.

The differences between the two delivery methods center on the data gathering methods. Collecting student reactions for Level 1 data and measuring knowledge or skills gained for Level 2, are easier to build into elearning courseware, which makes it easier to compile and analyze the data later.

On the other hand, data collection methods for remote learners often cannot include focus groups or direct observation due to logistical or budgetary reasons (ASTD, 2000). In addition, the high initial cost of either converting classroom training to elearning, or

starting a training capability by buying elearning courses, leads business leaders to demand better metrics than those provided in the past.

How Evaluation is Currently Conducted

For a variety of reasons, evaluation is often not conducted. The reasons for this range from the training budget being devoted to course development and designer salaries, with nothing left for evaluation, to management never having asked for any evaluation beyond course-completion rates. If training evaluation is done, it is often limited to Levels 1 and 2 because the trainers do not understand how to quantify the huge amounts of hard and soft data that a Learning Management System (LMS) can produce that is useful to derive on-the-job performance (Level 3) or business results (Level 4).

Organizations whose employees belong to unions have also been limited by union contractual constraints to use only Level 1 evaluations to prevent higher levels of evaluation from being used in performance reviews (Hall & LeCavalier, 2000). In a 2000 research study of 11 major corporations with significant elearning investments conducted, less than half were collecting business results data, although most managed to document Levels 1 and 2 (Hall & LeCavalier, 2000).

The 2002 ASTD SOIR found that only 33 percent of the companies surveyed try to measure learning effectiveness (Level 3) and that 12 percent try to measure job and business impact (Level 4) (Bersin, 2003).

The depth of the evaluation should depend on what companies really want to accomplish with it. There can be confusion on this point as well. For example, is the evaluation really measuring performance effectiveness (ensuring that your students can bake a Boston Crème Pie), or course completion (teaching X number of students to bake the pie).

Effectiveness is an objective comparison of actual results versus intended results. It is one of the factors that make up an evaluation. Evaluations must analyze many factors and derive meaning in order to guide decision-making. Of course, not every organization or course needs all levels of data. Here are some common reasons for doing a Level 4 or other results-oriented evaluation:

- **Identifying strengths and weaknesses of an elearning project**
- **Determining ROI**
- **Deciding who should participate in future elearning offerings**
- **Identifying who benefited the most or least from this course**
- **Collecting data to market future offerings**
- **Most important, building momentum for future elearning initiatives (Phillips et. al, 2002)**

WHY EVALUATE TRAINING

Historically, only the training departments had any interest in evaluation and their reasons were internally motivated. They did it to justify their existence by producing metrics of courses given and students taught, to decide whether to continue or eliminate courses, and to learn how to improve future offerings. For the most part, they measured activities not results. Trainers preferred Level 1 and 2 evaluations because they reflected the successes of the trainers. However, the last decade has shown a marked interest by leaders throughout the business world in making their employees more prepared for the rapid pace of changes in the global economy. Leaders see training, and especially elearning, as the best strategy for performance improvement.

They also intend to make trainers prove that the benefits of training outweigh the costs (Raths, 2001). In the best cases, the cost and benefits of converting classroom instruction to elearning have led to strategic, enterprise-wide imperatives binding training objectives

and business goals together. That means trainers should be routinely conducting Level 4 evaluations to demonstrate training's contributions. Still, that is not happening.

Why are Level 4s so hard to implement? Trainers believe that Level 4s are too difficult and costly because they misunderstand Kirkpatrick's label of "organizational results." There may be some misunderstanding that these organization results must equate to a financial figure such as sales, revenue or customer increases, fewer rework or produce defects, or more tasks completed. In other words, the myth is that organizational results require quantifiable, hard data. If the evaluation does not show positive results training will be blamed.

Another problem is that trainers believe they must be able to show proof beyond a reasonable doubt that business improvements were a *direct* result of a training intervention. This proof is often impossible to demonstrate.

Furthermore, proof could also be considered irrelevant, since in complex, interrelated, holistic systems that characterize organizations, rarely is any single cause, such as training, solely responsible for particular outcomes (Berge, 2004; Teletraining Institute, n.d.). What trainers need to remember is that they should show *evidence* that training contributed to a business result.

Level 4s are also much harder if trainers have not already conducted a Level 3 evaluation. Level 3 evaluations should provide evidence of the transfer of the new knowledge and skills to the workplace. If transfer did not happen, there are no Level 4 results to document. Finally, Level 4s are impossible if they are not appropriate for the type of training given. For example, if the course was designed to change attitudes, it may not produce observable or measurable outcomes.

THE PUSH FOR LEVEL 5 EVALUATIONS

Because elearning as a mode of delivery is much more accessible to all levels of employees in an organization than classroom instruction, it changes the importance and visibility of training. More scrutiny by senior leaders is often done and that is why more relevant evaluations become crucial.

Many business professionals are buying into the "continuous learning" potential of distance education, yet they see no value in doing Level 3 and 4 evaluations. "C level" executives want more concrete proof, usually meaning ROI.

The proof they are looking for takes the form of improvements in sales, productivity, quality, morale, turnover, safety records and profits (Kirkpatrick, 1998. However, Kirkpatrick's levels were designed for finite interventions, not continuous learning strategies. Therefore, in 1995 Jack Phillips developed the fifth level of evaluation—ROI.

This level strives to show the correlation between the money spent on the training and the monetary benefits produced. In fact, many companies involved in Level 5 evaluations now completely ignore Levels 1 and 2 because they do not contribute to measuring business objectives.

Still, ROI evaluations have both advantages and disadvantages that prudent evaluators must consider before jumping headlong into Level 5 (Adelgais, 2001). Additionally, the value of ROI differs from manager to manager and is not valued the same to all levels of an organization.

Table: 1
Differing levels of ROI use.

POSITION	GOAL	MEASUREMENT	SCOPE	PERSPECTIVE
Training Manager	Close skills gap	Individual performance; prefer their results via software they can drill down into for precise data	Business unit, specific training	Returns on training investments come from satisfying the needs of the Business Unit Managers—the only valid training ROI is business ROI
Business Unit Manager	Achieve business goal	Project goals, increased output, reduced absenteeism, increased employee morale and involvement, better educated workforce; prefer their results in spreadsheets and charts tailored to their unit		They own the problems that training solves—their question for training is, "What's in it for me?"
Senior Executives	Gain competitive advantage, transformation	Profit, cash flow, margin, stock price, venture capital; prefer results in charts or graphic displays	Enterprise, elearning infrastructure	Use strategy to create an environment where people learn faster and better than the competition

Adapted from Bersin, 2003

In many cases, trainers are attempting to serve the wrong stakeholders when it comes to conducting ROI evaluations by gearing their data-gathering for the senior executives' needs (see "measurement" in Table: 1).

However, the business unit managers are more interested in different measurements that are usually covered in Level 3 and 4 evaluations. At these stakeholder levels, the learners do not figure into ROI evaluations since their needs are covered by Level 1 and 2 evaluations (ASTD, 2003). Another problem trainers inflict on themselves is providing the wrong ROI metrics because they do not understand the business problem their courses are supposed to address.

Trainers must partner with the business units to learn what metrics each unit values, what are its future strategies, pitfalls, and what quantifiable outcomes are desired (Purcell, 2000).

Converting business results to monetary values is the thorny problem that prevents many organizations from even attempting to claim ROI success. The general steps for Phillips' model (Phillips, 1996) are simple, but the details and implementation can be difficult:

- **Step 1: Collect Level 4 evaluation data. Getting to Level 5 requires that *there is already preliminary results from the previous levels*. Determining business results (Level 4) implies that they have already found improvement based on the learning (Level 2), and measured application of new skills on the job (Level 3). This step takes trainers with measurement skills, which most training activities do not have.**
- **Step 2: Isolate the effects of training from the other factors surrounding the business results.**
  There are many ways to do this but using a control group is the most common, as long as you also have a pre-training history of production output and can follow-up on the test and control groups for 60-90 days to compare post-training performance.
- **Step 3: Convert the results to monetary benefits. Trainers must separate their results into hard and soft data. Hard data: *output* (units produce, forms processed, tasks completed), *quality* (scrap produced, waste eliminated, product rejects), *time* (equipment downtime, employee overtime, training time), or *cost* (accident costs, sales expenses, overhead). Soft data: examples include *work climate* (grievances, turnover), *work habits* (absenteeism, tardiness), *new skills* (decisions made, problems solved), and *initiative* (follow-through on new ideas, project completion, employee suggestions). This is a very subjective process, and even Phillips warns that not all data can or should be converted to monetary values, since intangible outcomes help create a balanced assessment of the results.**
- **Step 4: Total the training costs: Designer and programmer salaries, development and implementation costs, marketing and any other costs directly related to the course.**
- **Step 5: Compare the monetary benefits from Step 3 to the training costs in Step 4. Did the benefits exceed the costs?**

Although time-consuming, it is often possible to calculate ROI and stick with the modified five-levels model. Some companies forego the complex statistics by limiting their ROI analysis to their most critical measures. However, other professionals discount measuring ROI altogether because they think this financial-only outlook is too short-sighted. They see ROI as a snapshot in time, a lagging indicator of where the organization was, but not where it is going or how best to get there. Instead, they are devising their own measurement devices.

FUTURE TRENDS

In 1993 Kaplan and Norton wrote that financial metrics were out of sync with the skills and competencies that companies now measure, so they created the Balanced Scorecard (Abernathy, 1999). This model was designed to measure and manage performance throughout the entire organization, not just training. It is important to understand the big picture before looking at how it relates to training. The Scorecard measures performance across four key perspectives:

- **Financial (revenue growth, cost management, asset utilization)**
- **Customer (market share, customer retention/acquisition/satisfaction/profitability)**
- **Internal business processes (identify the market, design/build/deliver products or services, after-sales service)**
- **Learning and growth (employee capabilities [where training initially fits], motivation, information technology capabilities)**

Unlike ROI which only reports on the past, the Scorecard sharpens an organization's focus on future success by setting and balancing objectives for each of these perspectives, creating drivers of future financial performance (Willyerd, 1997). Each business unit devises vision statements at their level for each of those perspectives, defines its critical

success factors for each, and then identifies the specific measurements that will decide whether the factors were satisfied. Where does training come in? Trainers ensure the courses are aligned with each business unit's critical success factors and measurements, if possible, from all four perspectives. Training can directly affect the "learning and growth" perspective of nearly all business units if it is understood what the units need. It can indirectly affect the other three as well, so evaluations are critical to satisfying the business unit's goals. The Scorecard requires every action to show accountability to established corporate goals, thereby promoting efficient activity alignment and eliminates projects that do not contribute to strategic success (Willyerd, 1997).

Not everything needs to be measured. Results depend on *what management thinks is important*. For example, Hodges measures return on expectation (ROE) (Goldwasser, 2001). After finding that Verizon's business units did not track the data she needed for Level 4 and 5 evaluations, it was impossible to isolate the specific effects of training. So she devised a way to measure the senior leaders' expectations for the training. A trained facilitator conduct an interview (15-20 minutes) with a key executive in the business unit to pin down learning objectives and identify what will constitute proof of success. Once the training is finished, the executive is interviewed again and asked to quantify the results and put a monetary value on the change. This constitutes reasonable evidence for ROI calculations. She notes that it takes a skilled interviewer to both extract and quantify assessment information. When she has been able to conduct corresponding ROI evaluations, those results match the ROE results every time.

The "time-to-competency" model (Raths, 2001) is often used in call-center training, when it takes at least three months for operators to achieve competency after classroom training. Unfortunately, turnover is high and trained employees would often leave the company after a year. Time to competency starts with establishing a baseline of the skills and production available at the beginning of a new elearning initiative. After training is completed and a database of frequently asked questions went online, the skills and production at measured at 3, 6 and 12-month intervals to determine competency rates and track employee turnover. This satisfied management's desire to quantify training efficiency. Another example measures elearning's return and calls it "time-to-market" (Raths, 2001). Software developers take six to eight weeks to train their sales staff on new products before they can release the product. But by the time they are up to speed, the software is into another revision. They now use elearning to provide the sales staff chunks of information as the product is developed, so that they are ready to go when it is released, saving four to six weeks on their competitors.

Christine Pope, director of elearning services for SmartForce, teaches customers to take a three-pronged approach to evaluation: first look at cost savings (of elearning results over classroom results), then move to performance improvement (involving supervisory evaluations and financial data beyond training metrics), finishing with competitive advantage, or bottom-line results (Raths, 2001). LMS help in this regard by both providing the courses and gathering data about their usage.

Finally, there is a growing number of training professionals that think after-the-fact evaluations are entirely the wrong way to go—they are promoting ROI forecasting. Still, they see clear links between ROI evaluations and forecasting (Graber, n.d.). Forecasts rely on accurate evaluations of training costs and impact, while evaluations benefit from the baseline data that forecasts can supply. This method was implemented at Commonwealth Edison, where they identified several advantages of forecasting:

- **Forecasts identify the highest ROI alternatives, by comparing expected values of a range of training options and choosing the highest outcome. Evaluations after training cannot tell you if another option would have been better.**
- **It helps avoid poor-but-costly investments: the most useful information is that which helps make the *investment* decision. Determining ROI after the training is only justification.**

- **It is consistent with standard business practice: analysis *before* the decision is the sound approach, while putting money and time into analyzing training already completed is less appealing to business.**
- **Forecasts are much easier to produce than ROI evaluations, especially when a pilot program may not be possible.**
- **ROI forecasts may be used to justify training's value against competing budget needs.**

CONCLUSION

Measuring for Level 4 or 5 has always been very difficult for trainers, since they usually do not have the staff time, budget and expertise necessary. The high costs of elearning and the added versatility of learning management systems to capture data have combined to interest business leaders in forcing training to improve addressing and measuring business results. Kirkpatrick's four levels were not designed to produce the ROI, ROE or other metrics that the business world now demands. Phillips' Level 5 does provide a manageable, albeit laborious, way to satisfy those requirements. Other training professionals are inventing new ways to evaluate training and correlate its results to business objectives. Regardless of what is measured or how, the consensus seems to be that what is important is that business values are finally being attached to the corporate learning experience. Holly Burkett, an ROI evaluator at Apple Computer, stated "For me, it's more empowering to know that our department's work has a direct impact on performance, productivity, or sales than it is to know that people enjoy the training program" (Purcell, 2000).

BIODATA AND CONTACT ADDRESSES OF AUTHORS

Zane L. BERGE is Associate Professor of Education. Ph.D., 1988, Michigan State University, Educational Systems Development (Dissertation: Effects of group size, gender, and ability grouping on learning science process skills using microcomputers.). B.S., 1977, Rochester Institute of Technology, Photographic Finishing and Production ManagementCertification in Distance Education, 1995, Pennsylvania State University. Honors Received:
1999-Charles A. Wedemeyer Award for distinguished scholarship and publication (Distance Training)
1987 Finalist - Spencer Dissertation Fellowship Competition
1985-present. Phi Kappa Phi National Honor Society, Life Member
1983-1986. Fellowship, Institute for Research on Teaching, Michigan State University

Associate Professor Zane Berge, Ph.D.
9525 Pamplona Road,
Columbia, MD 21045 - 410-455-2306
Email: berge@umbc.edu and http:// www.emoderators.com

REFERENCES

Abernathy, D. J. (1999). *Thinking outside the evaluation box.* Retrieved July 18, 2004 from http://www.pdonline.ascd.org/pd_html/eval2read1.html.

Adelgais, S. (2001). *Return on investment—An evaluative framework.* Retrieved July 18, 2004 from http://coe.sdsu.edu/eet/Articles/roi/index.htm.

Berge, Z.L. (2004). Complexity and confusion in distance education. *Distance Learning.* 1(2): 1-6.

Bersin, J. (2003). *Elearning analytics*. Retrieved July 18, 2004 from http://www.learningcircuits.org/2003/jun2003/bersin.html.

Goldwasser, D. (2001). *Beyond ROI.* Retrieved July 18, 2004 from http://www.trainingmag.com/training/search/search_display.jsp?vnu_content_id=1504012

Graber, J., Post, G., & Erwin, R. (n.d.). *Using ROI forecasting to develop a high-impact, high-volume training curriculum.* Retrieved July 18, 2004 from http://www.businessdecisions.com/Docs/ASTD ROI Chapter.doc.

Hall, B., & LeCavalier, J. (2000). *The case for level 3*. Retrieved July 18, 2004 from http://www.learningcircuits.org/nov2000/hall.html.

Kirkpatrick, D. L. (1998). *Evaluating training programs: The four levels*, 2nd Edition. San Francisco: Berrett-Koehler Publishers, Inc.

Phillips, J. J. (1996, March). Was it the Training? *Training and Development*, pp. 28-32.

Phillips, J. J., Phillips, P. P., Duresky, L. Z., & Gaudet, C. (2002). Evaluating the return on investment of elearning. In Alison Rossett (Ed.), *The ASTD Elearning Handbook* (pp.387-397). Madison, WI: The McGraw-Hill Companies.

Purcell, A. (2000). *20/20 ROI*. Retrieved July 18, 2004 from http://www.astd.org/members/td_magazine/td0700/purcell/0700.pdf.

Raths, D. (2001, May). Measure of success. *Online Learning*, *5*(5), 20-22, & 24.

Rothwell, W.J., Sanders, E.S., & Soper, J.G. (1999). *ASTD models for workplace learning & performance.* American Society for Training & Development.

Teletraining Institute (n.d.). *Embracing evaluation.* Retrieved November 27, 2003 from http://www.teletraining.com/LP/embracing_evaluation.htm.

Willyerd, K. A. (1997). Balancing your evaluation Act. In Donald Kirkpatrick, *Evaluating Training Programs: The Four Levels,* (2nd Ed.) (pp.87-97). San Francisco: Berrett-Koehler Publishers, Inc.