

Are You Too Nice to Train? (Aug 06)

By Sarah Boehle

A little evaluation can be a dangerous thing. Just ask Neil Rackham. Years ago, the best-selling author of SPIN Selling and Major Account Sales Strategy was asked by a European technology company to examine the way it evaluated trainer performance. "They were using Level I methods at the end of each program, which consisted of giving out a questionnaire to students at the end of training that basically asked, "How do you feel about the trainer?" and "Do you think he or she was effective?"

Two trainers in particular consistently received poor ratings. As one might expect, management began to wonder aloud about those trainers' futures. "One of the trainers had applied for a management position, and managers were wondering whether they should even consider him for a promotion if his evaluations didn't seem to be any good, and whether consistently high evaluation scores from students should be a qualification for moving to the next level."

Rackham decided to dig deeper. The results of his research were startling, to say the least. Turns out, the two most abysmally rated trainers in the company were actually the best in their quartiles-and often the best on staff-when it came to learning gains for their students. "In the end," Rackham says, "Level I smile sheets had given management the exact wrong impression."

If you think Rackham's story is an anomaly in the training biz, consider the case of Century 21 Real Estate. When Roger Chevalier joined the organization as vice president of performance in 1995, the company trained approximately 20,000 new agents annually using more than 100 trainers in various U.S. locations. At the time, the real estate giant's only methods of evaluating this training's effectiveness-and of trainer performance, for that matter-were Level I smile sheets and Level II pre- and post-tests. When Chevalier assumed his role with the company, he was informed that a number of instructors were suspect based on Level I and II student feedback.

Chevalier set out to change the system. His team tracked graduates of each course based on number of listings, sales and commissions generated post-training (Level IV). These numbers were then cross-referenced to the office where the agents worked and the instructor who delivered their training. What did he find? A Century 21 trainer with some of the lowest Level I scores was responsible for the highest performance outcomes post-training, as measured by his graduates' productivity. That trainer, who was rated in the bottom third of all trainers by his students in Level I satisfaction evaluations, was found to be one of the most effective in terms of how his students performed during the first three months after they graduated.

"There turned out to be very little correlation between Level I evaluations and how well people actually did when they reached the field," says Chevalier, now an independent performance consultant in California. "The problem is not with doing Level I and II evaluations; the problem is that too many organizations make decisions without the benefit of Level III and IV results."

Just how common is it for Level I results to give management the wrong impression? According to the research, very.

Will Thalheimer of Work-Learning Research, a research-based consulting firm in Somerville, Mass., points to a study by George Alliger entitled "A Meta-Analysis of the Relations Among Training Criteria" (Personnel Psychology, 1997), the results of which indicate that there is an exceedingly weak correlation among the various levels of training evaluation. The conclusion? Thalheimer says: "Just because a trainer receives a relatively good rating on his or her Level I evaluations, that's telling us very little about how much learning is taking place or how much on-the-job performance is actually impacted," Thalheimer says. "High smile-sheet ratings translate neither to learning nor to high performance on the job."

In some instances, there is not only a low correlation between Level I and subsequent levels of evaluation, but a negative one. Richard E. Clark, a professor of educational psychology and technology at the Rossier School of Education at the University of Southern California, has done rigorous research involving the effectiveness of Level I as an evaluation tool. His findings, published in a 2002 book, *Turning Research into Results: A Guide to Selecting the Right Performance Solutions*, reveal that there is a negative correlation between Level I results and on-the-job performance (Level III). "When used to evaluate performance improvement programs, reaction questionnaires, or 'smile sheets,' often indicate the opposite of what happened," say Clark and his co-author Fred Estes-either rating an effective course poorly or an ineffective course highly.

Why Smiley Sheets Stink

Research has established the fact that smile-sheet results are an extremely poor indicator of how much students actually learn or how well they are likely to perform post-training, but the tool's drawbacks don't stop there.

Smile sheets require learners to judge the quality of their own learning experiences. The problem with this practice, Clark says, is that typical Level I reaction evaluation only indicates whether people liked something-and not whether they learned anything at all. Learners also are "famously optimistic" about what they will remember, Thalheimer says.

"In some cases," Clark says, "people like training courses where they learn almost nothing. In fact, there is evidence that some people leave training courses knowing significantly less about something than they did when they started the course-and yet they sometimes like the 'unlearning' experience. Similarly, negative reactions do not always indicate that people have learned a great deal. But in a significant number of cases ... people report disliking training where they have gained a great deal. The fact is that people's preferences-their likes and dislikes for training-are totally unreliable indicators of what they've learned."

Rackham, who has conducted a good deal of research in this area, concurs. "The fact is, people are very poor judges when it comes to perceiving when they are learning. They are generally much better at perceiving when they are not learning," he says. "If people are having a good time, they will very often perceive their learning to be more than it actually is. But high enjoyment is not necessarily related to high learning. If you have a trainer who tells 100 war stories and is very entertaining, that instructor can end up getting tremendous ratings for 'perceived learning,' but two hours later, trainees can't remember a single thing that came out of the session. The danger is that we have some great entertainers in the training ranks who may rate very high on enjoyment but are so busy entertaining the class in order to get good ratings that they're delivering pitiful learning."

Our Own Worst Enemy

Problem is, far too many organizations use Level I not as an unscientific qualitative indicator, but as a comprehensive instrument for determining everything related to training. These organizations over-rely on Level Is for purposes they never were meant to serve, such as determining instructor performance, business outcomes and learning and course effectiveness.

Clark's take on the reason for this phenomenon is that some people "hate to invest the kind of mental effort required to learn something complex, while others love it and a few can take it or leave it." He also suspects that "most of the managers who look at the reaction data think they tell them something useful about learning and job performance."

In higher ed, in particular, research points to the danger of Level I in this regard. The Ivory Tower is infamous for tying a large portion of professors' upward mobility to student evaluations. These evaluations

often penalize those professors who challenge students (and grade accordingly) and reward professors who take it easy on students and are likeable and charismatic.

Thankfully, performance evaluation systems are not quite as rigid in the for-profit world- but abuse still exists. "Not many people are really concerned openly about the Level I issue in corporate America," says Jim Kirkpatrick, a senior consultant in the evaluation practice area at Corporate University Enterprise, "but it is an absolutely critical issue. I think there is an element that sneaks over from higher ed in that there are still a lot of more traditional companies that evaluate trainers on a five-point scale based on Level I smile sheets. In these organizations, questions on the questionnaires typically are not worded appropriately, the scores are misused, and they are most certainly tied into trainers' performance review and merit pay."

Some training organizations may be apt to pooh-pooh this problem by arguing that they are smart about using Level I evaluations, that they take the results they glean from them with a grain of salt, and that they realize that Level Is are but one source of data among many. While that may indeed be the case within more progressive organizations, don't be so sure that such "best practices" extend to the majority of U.S. organizations.

The American Society of Training & Development's (ASTD) 2005 State of the Industry Report provided the results of a benchmarking survey of large global organizations, most of which are based in the U.S. In the survey, these organizations were asked whether they evaluated training at each of the four levels (based on Donald Kirkpatrick's four-step model).

Ninety-one point three percent reported evaluating their training programs at Level I, and 53.9 percent reported evaluating at Level II. Meanwhile, only 22.9 percent of organizations surveyed reported evaluating at Level III, and only 7.4 percent evaluated at Level IV.

Chevalier estimates that even these numbers are inflated for the training industry as a whole because the organizations represented voluntarily participate in the study each year and are a sample of convenience. "My best guess, based on personal experience and conversations with colleagues, is that the number legitimately doing Level IV evaluation is closer to two percent nationally."

Regardless of whether actual percentages are indeed lower than those reported, the picture ASTD's findings paint isn't a pretty one. They suggest that a large number of training organizations are doing either no evaluation at all or Level I evaluation only.

And if Level I is virtually the only data point that many companies are using to determine the success of both their trainers and their training, it doesn't take a large leap of logic to surmise the type of data on which they rely to justify their existence.

"We are our own worst enemy," says Mark Hilldrup, who works in the evaluation practice area at Corporate University Enterprise in Falls Church, Va. "In the absence of buy-in or approval to go beyond Level I to do other levels of evaluation, Level I is all that trainers really have to show for their work. As a result, what often happens is that the first thing trainers report on their scorecard is their Level I aggregate each month-so, we as a profession turn around and promote the same dysfunctional thinking about Level I by treating it as a very important thing."

This practice can be especially dangerous when evaluation forms aren't customized for each class, and when qualitative data gleaned from Level Is are used in a quantitative manner, Hilldrup notes. "The complaint you often hear from trainers is that only four out of 15 questions on an evaluation typically have anything to do with them. They aren't related to content; they're related to the facility or the food-things that trainers have no control over. Then, somehow, all those numbers get rolled up into an overall figure—a benchmark, if you will -and shown around the office, and it doesn't take long for people to figure out who the 'good' trainers are."

Good Evaluation: Why Bother?

Why do so many companies make this mistake? In all likelihood, they do so because the steps they must take in order to get their hands on the objective information necessary to accurately assess training effectiveness can be complex, difficult, costly and time-intensive.

Smile sheets, however, are easy. They don't require a lot of time or money to administer, and they offer a way to churn out results without a lot of pain. Additionally, learners and their managers expect them.

Yet another reason training departments may not go beyond Level I is that their clients aren't demanding it. A joint ASTD/IBM study, "C-Level Perceptions of the Strategic Value of Learning," published in January, compared CLOs' perceptions of the value of learning with CXOs' perceptions. Results indicated that CXOs (non-learning executives such as CEOs and CFOs) were very comfortable getting Level I responses only. CXOs, the report also claimed, "were more interested in perceptions of value and alignment of learning with business needs than in quantitative data to prove the value of training."

"In other words," Thalheimer says, "if courses were well received by learners and employees, CXOs were happy with that. What does that tell you? Smart training managers and CLOs say, 'That's what we need to focus on. We need to make sure our Level I scores are good and our courses are well received because that's what the CEO is asking for.' It also tells us that CLOs are doing a lousy job of educating their bosses around the value of good assessment."

Finally, there's the proverbial elephant in the room: Trainers may refrain from doing subsequent levels of evaluation out of fear of what they'll find. Allen Interactions once performed an interesting-though unscientific-straw poll. For two years, the Minneapolis e-learning custom content firm, in an effort to promote evaluation use, offered clients its services for 5 percent off for those who agreed to evaluate their e-learning program, and to sweeten the pot, threw in an evaluation of clients' training at half cost-offering the equivalent of approximately \$100,000 worth of evaluation for free. "What percentage of Allen Interactions clients do you think took them up on this?" Thalheimer asks. "The answer is zero."

What gives? "From my experience talking to corporate training departments, the calculus in the field seems to be, 'Don't do good evaluations,' "Thalheimer says. "Why would anyone want to? If you evaluate at all levels, only bad things can happen. If you use Level I and get good results, then you're maintaining the status quo and everyone assumes that training is doing a good job. If you evaluate more rigorously and get bad results, however, all hell breaks loose."

That kind of thinking is reprehensible, Thalheimer notes, because training departments have a professional, "perhaps moral," responsibility to do fair and valid assessments. "Most current practices are shameful. We get some feedback, but we don't get the valid feedback we need to hold ourselves accountable."

Nevertheless, training departments aren't apt to change their ways any time soon, he says-unless they and those to whom they report finally decide to get serious about doing authentic evaluation.

Until that happens-or rather, if it ever happens-proceed with caution when using Level I. If you don't, someone-be it your trainers, your learners, or even your business-might suffer as a result.

Training Magazine, ©2006 VNU Business Media