New Routes to Evaluation

BY NANCY M. DIXON

What's new in evaluation? If you talk to leading-edge companies, quite a bit. Take a look at five top companies to discover the latest trends, including new roles for customers and less use of evaluation to justify training.

AFTER YEARS of saying, "we know we ought to evaluate but..." some companies really are changing the face of training evaluation. These new methods, however, may not be what you would expect.

Through a study of more than 40 organizations, I identified IBM, Motorola University, Arthur Andersen, Florida Power and Light's Nuclear Division, and the AT&T School of Business as the companies with the best training evaluation practices. Here is what made these companies stand out:

- measuring customer requirements
- testing participants
- measuring what the client can use and will pay for
- moving away from justification.

Customer focus

Traditionally, Level I evaluation has gathered participant reactions at the end of a program. However, in these organizations, Level I consists of customer satisfaction indicators. In the past, trainers designed participant reaction forms — often using the best ideas they could garner from their collection of forms. In best-practice companies, the customer identifies what is measured. Switching from trainer focus to customer focus requires different thinking about the purpose of evaluation. The first step is to identify customer requirements.

In these companies the identification of customer requirements is detailed and extensive and the methods to collect the information varied and multiple. IBM uses focus groups and interviews. Arthur Andersen uses focus groups and concept mapping with customer groups. This type of data collection is far more comprehensive than traditional needs assessment. In these companies, customer requirements involve every aspect of the interaction from the time a customer call begins until results are achieved.

Trainers are often convinced that they know what customers want—they often hear managers say, "We need more how-to courses but held over shorter time frames." Yet, as Barry Arnett of IBM notes, well over half of what IBM identified as customer requirements had not been previously measured by the training unit. Like IBM, the best-practice companies that have investigated customer requirements are often surprised by what is important to their customers.

Additionally, the meaning of the term "customer" has expanded in these companies to include participants, their managers, external customers, training councils, vendors, regulatory agencies, and upper-level management.

Meaningful standards

The second step in changing the focus to customers is to translate customer requirements into meaningful and useful standards. For example, if customers want their service people trained before product release, then a standard might be established to offer training a minimum of two months before the release date of any product. Or, if customers want employees certified in certain tasks, then a standard might be established to have 90 percent of participants pass the certification examination on the first administration.

Turning requirements into standards is done in collaboration with customers; it is their standards that trainers are trying to meet and therefore their involvement is critical. It is important to note that identifying customer requirements is not the same thing as giving customers everything they demand. It takes considerably more digging and analysis to move beyond requests and complaints to truly understand customer requirements and to establish meaningful standards.

The third step is to translate the standards into indicators that, over time, can measure improvement. In the examination example above, data collection might simply involve designing a process to consistently collect test scores. In the other situations, surveys or questionnaires might need to be designed to collect the data. In still others, the customer might already own useful data and might be able to provide it to the training unit.

The number of indicators for meeting customer requirements varied among the best-practice companies; IBM has developed 16, AT&T has six, while FPL Nuclear has indicators in 12 categories. At FPL Nuclear, each indicator is assigned to an individual who is responsible for updating and displaying that indicator so everyone sees how well the system is meeting customer requirements. People who've used this practice say that indicators become more useful as they are refined over time.

The fourth step in changing the focus to customers is to improve the scores on the indicators. That involves the training unit focusing on its own processes—using the data to improve how things are done. In addition to improving the scores on the indicators, IBM says it decreased the cost of training by over 200 million dollars a year in its United States' organizations. FPL Nuclear reduced its cost by almost one half, while providing the same amount of training to essentially the same number of participants.

Training requires assistance from other parts of the organization to meet customer requirements fully. For example, participants may not be able to implement what they learn in a program because the tools they use on the job do not support the learning, or because the measurement system may encourage employees to do the opposite of what they learned.

The fifth step is communicating indicators to customer groups. Throughout the entire process, training should cooperate with customers to:

- identify requirements
- set standards
- assist in designing indicators that have meaning to customers
- assist in collecting data
- help resolve root causes of problems.

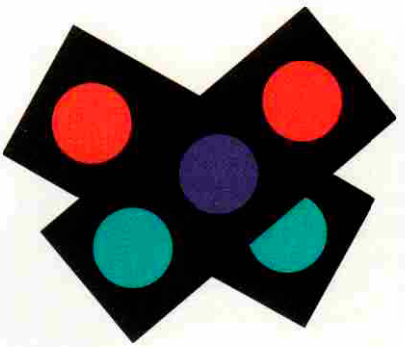
According to these companies, it is critical to inform customers about the efforts the training department is making in meeting customer requirements. This demonstrates value and illustrates that the training department is responsive to its customers. At IBM and FPL Nuclear, these efforts have increased customer loyalty and customer satisfaction.

Meeting customer requirements is a continuous activity rather than a onetime event. These companies already have processes to update customer requirements, raise standards, develop better indicators, and improve their own practices to meet customer requirements better.

Just testing

In the companies I studied, the meaning and use of Level II evaluation — testing or demonstrating performance— has also broadened in scope from the conventional interpretation. Practice and policies varied widely across the organizations. IBM found that when they implemented an effective instructional system design process, much of their testing became redundant. This allowed the company to refocus their evaluation efforts on behavior change and impact.

Conversely, at Motorola, the use of tests is on the increase. Currently, about one-third of their courses include some form of testing, but their plan is to include tests in 100 percent of courses in the future. Tests serve two purposes for Motorola. During the pilot phase of development, tests provide vital information for course



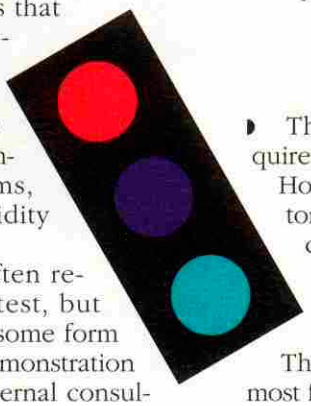
improvement. Motorola's expanded testing also provides feedback to participants about their learning. This feedback is coupled with remediation when participants score below an acceptable level. At AT&T, testing is used only when it enhances learning, for example, if pre-tests serve as a useful advanced organizer or if post-tests reinforce content.

At FPL Nuclear, testing serves three purposes. On the front end, testing is used regularly as an analysis tool to determine the level of training needed. Tests administered at the end of training affirm that participants are qualified to do critical tasks—an important consideration in a heavily regulated industry. Finally, when administered months after training, retention tests determine when and what kind of refresher training is necessary.

By contrast, Arthur Andersen uses both pre- and post-testing, but only in courses that have a knowledge base. Practical application courses do not have tests. AT&T uses e-mail to collect pre- and post-test data on many of its courses, while companies such as Motorola prefer to embed tests throughout courses rather than waiting until the end. These examples

illustrate that the need and use of tests has not been standardized across these best-practice companies.

However, there are some trends. One critical trend is the increased use of certification and the testing that it necessitates. Certification, as the verification of employee competence for a specific job function, is in much wider use, perhaps spurred by ISO 9001. The interest in certification appears also to encompass less technical jobs such as service representative or internal consultant. To meet the certification demand, instructional system designers are sharpening their knowledge and skills in test development to ensure that they create legally defensible certifications that have job performance validity. In some companies certification is being linked to compensation systems, again making validity critical.



Certification often requires a written test, but most also require some form of performance demonstration or project. IBM internal consultants must pass written tests, but an important part of their certification is to engage in a major consulting project and then produce evidence of their competence for a certification board. The cost of developing valid and reliable written tests and appropriate performance criteria and demonstrations is very high.

The companies are becoming much more skillful in the development of tests and performance demonstrations, yet they are also finding the cost and time involved are too great unless the test does more than determine if participants have learned. When the test can be used as a way to enhance learning, as a basis for compensation, or as a product itself, then the cost is more justifiable.

What the client needs — and will pay for

All of the best-practice companies are conducting Level III (change in behavior) and IV (impact on the organization) evaluation selectively rather than consistently. This appears to be a conscious strategy on the part of these organizations rather than a default position. The strategy is based on three concerns that are well known to trainers:

- Data collection at this level has to be customized for each situation — there are no generic forms or processes that can capture impact data.
- Data collection is costly and time consuming.

> ■ *Certification is sometimes linked to compensation, making validity critical* ■

- This kind of data collection requires the collaboration of clients.

However, there are two critical factors that prompt the companies to collect evaluation data: the client feels a need for information, and the client is willing to pay for data collection and analysis. These two factors come together most frequently when a client requests training's help in addressing a specific problem or issues within a unit.

The trainers in these companies are often similar to consultants. The training unit works with the clients to assess the problem, and contracts with them for needed changes in management practices, compensation, and training. As part of the initial consultation, the client and the trainer identify measures that will have meaning to the client (and also can serve as evaluation for the training department) and plan how the results data will be obtained. Existing business measures are most frequently used under these circumstances, rather than newly collected data that would only serve as evaluation measures.

Darrell Jinkerson, at Arthur Andersen, explains that it is important for trainers to become familiar with the data that already exists within a system so it can be used when the opportunity arises. Jo Magennis, at FPL Nuclear, notes that "providing value-added training requires a partnership between line management and training. This partnership has three critical components: line customers who identify performance problems that require training solutions, a training organization that responds with training that specifically addresses the skill and knowledge deficiencies and does so in a timely manner, and indicators of measurable improvement in post-training performance."

Desirable results

When an intervention involves several changes in addition to training, it is not possible to provide proof that results were based solely on the training effort. However, these best-practice companies seem less concerned with collecting irrefutable evidence than with determining whether the desired result was obtained. In some special situations, training has been able to identify control groups or to account for other variables in ways that would satisfy even the most critical statistician.

Trainers in these organizations actively encourage their clients to collect results data. Although not always persuasive in their efforts, trainers recognize that any successful intervention requires measured results. Moreover, although these measures are collected for the benefit of the client, they are used by training. At Arthur Andersen, Level III and IV are measured less than 10 percent of the time, yet the studies that are conducted have such convincing results that they serve to greatly enhance training credibility throughout the organization. Such studies are very useful with future clients, but the benefits seem to spill over into open-enrollment programs. AT&T, like many of the best-practice companies, has been deliberate about packaging and distributing evaluation results to increase customer confidence in its product and service. Arthur Andersen thinks of these as landmark studies and actively promotes the results both internally and externally.

One of the most striking findings of this study is that none of the best practice organizations is evaluating primarily to justify training or to maintain the training budget. That appears to be true across all levels of evaluation as well as across the five companies. Level I, reorganized as customer

satisfaction, is being used to make training more effective in meeting customer requirements; Level II either is used to enhance learning or as a product itself; and Level III and IV are being used selectively when the evaluation data is useful to the client. This change in thinking is noteworthy because much of the training evaluation literature cautions trainers to collect data that proves to top management that training provides an appropriate return on investment. It seems almost paradoxical that when evaluation is finally implemented, it is primarily in the interest of a business need rather than a training need.

Most of these evaluation efforts have taken place within the past two to three years. Yet, in this brief time, they do appear to have evolved. In each of these organizations, when a number of evaluation processes have been attempted and found wanting, other practices are tried until a useful fit is found. While there are commonalities among the best-practice companies, the differences are striking, as well. Each evaluation system seems to be tailored to that particular organization, representing its culture and needs.

Each benchmark organization has an individual or team assigned to evaluation. However, these professionals do not necessarily conduct evaluation themselves; rather, they are focused on the development of a rationale and guidelines for evaluation and for the development of the evaluation process. Many of these practices, although not all, reside in a corporate or central unit rather than being consistent throughout these very large organizations. This may be, in part, because evaluation professionals tend to be assigned to corporate offices.

The trend in many organizations is toward outsourcing parts of the training. If we take analysis, design, development, delivery, and evaluation as the major elements of training, increasingly these best-practice organizations are outsourcing the delivery. For several years, Motorola has contracted out its delivery and IBM outsources well over 50 percent of its delivery — a number it intends to increase. Design and development are also increasingly outsourced in many of these organizations. However, the two pieces that are consistently retained in-house are analysis and evaluation. As evaluation is redefined in these organizations, it is easy to understand why it is vital that it be retained in-house. Evaluation is finally seen as critical — but not the way you might have expected. ■

Nancy M. Dixon *is an associate professor at George Washington University, Administrative Sciences Program, 2136 Pennsylvania Ave., NW, Suite 301, Washington, D.C. 20052; 202/496-8380.*