

1

Full-Scope Evaluation: Raising the Bar

SEELS AND RICHEY (1994, p. 52) call evaluation “a commonplace human activity” and indicate that as far back as the 1930s instructional designers, evaluators, and other training/HPT (human performance technology) practitioners discussed, wrote about, and sometimes implemented evaluation activities to measure the value of training and learning. The evaluation bar was raised in 1967 when Scriven suggested that exemplary instructional designers and evaluators plan and conduct two types of evaluation: *formative evaluation*, to improve instructional programs or products during the development phase; and *summative evaluation*, to measure the effectiveness of education, training, and learning during or immediately after implementation. The terms *formative* and *summative* have “not only served the field well in providing a usable language to describe important uses of evaluation, but have also been a rich conceptual seedbed for the sprouting of many proposed refinements and extensions to the field” (Worthen, Sanders, and Fitzpatrick, 1997, p. 18). Now it’s time to raise the bar again.

We challenge evaluation, training, and HPT practitioners to add confirmative evaluation to their repertoire of knowledge and skills. Confirmative evaluation goes beyond formative and summative evaluation to judge the continuing merit, value, or worth of a long-term training program. More specifically, we challenge training and evaluation practitioners to consistently use full-scope evaluation: formative, summative, confirmative, and meta. Confirmative evaluation encourages and supports continuous improvement efforts within organizations. Meta evaluation evaluates evaluation and adds credibility to evaluation activities. However, meta evaluation is another story and another book. Meanwhile, we need to focus on confirmative evaluation.

In this chapter, we set the stage for confirmative evaluation. First, we introduce the concept of full-scope evaluation as an integrated plan that uses four types of evaluation—formative, summative, confirmative, and meta—to judge the continuing merit and worth of long-term training programs. We use models to illustrate how the four types of evaluation work together and how full-scope evaluation fits into the instructional system design (ISD) process. Then we discuss the challenges faced by individuals and organizations that commit to full-scope evaluation.

One issue that arose when we began writing this book is that although there is common evaluation vocabulary, there is limited shared meaning. When discussing evaluation, the literature uses the words *types*, *roles*, *stages*, *phases*, and *forms* of evaluation. For consistency, we use the word *type* when referring to formative, summative, and confirmative evaluation.

After reading this chapter, you will be able to:

1. Explain the concept of full-scope evaluation
2. Describe and compare the components of full-scope evaluation (formative, summative, confirmative, and meta evaluation)
3. Explain how full-scope evaluation turns ADDIE into ADDI/E (more on this later; also, see the Glossary at the end of the book)
4. Recognize the challenges associated with committing to full-scope evaluation

Evaluation: The Full Scope

Full-scope evaluation systematically judges the merit and worth of a long-term training program before, during, and after implementation. Full-scope evaluation is appropriate only for training programs that are designed to run for one year or more; it is not appropriate for a one-time training event, such as a single-session workshop to introduce a new product to sales representatives.

Full-scope evaluation integrates four types of program evaluation—formative, summative, confirmative, and meta—into the training program evaluation plan (see Chapter Three). Working together, the four types of evaluation help to determine the value of a long-term training program and develop the business case or rationale for maintaining, changing, discarding, or replacing the program. We describe all four types of evaluation here.

Formative Evaluation

Formative evaluation is the oldest type of evaluation. Scriven (1967) was the first to use the term; however, the concept and practice of evaluating instruction during development predated both the term and the ISD movement (Tessmer, 1994). Thiagarajan (1991) defines and describes formative evaluation from a quality perspective as “a quality control method to improve, not prove, instructional effectiveness” (p. 22) and “a continuous process incorporated into different stages of development” (p. 26). Dick and King (1994) add that formative evaluation is a way to “. . . facilitate the transfer of learning from the classroom to the performance context” (p. 8).

Formative evaluation is usually conducted by the designer or developer; however, large organizations sometimes call on the services of a practitioner evaluator. Van Tiem, Moseley, and Dessinger (2000) describe four basic strategies for conducting formative evaluation:

1. Expert review using an individual or group familiar with the content and need
2. One-to-one evaluation involving the designer or evaluator and a learner or performer

3. Live or virtual small-group evaluation
4. Field testing or piloting either segments or all of the program or product (pp. 164–167)

The outputs and outcomes of formative evaluation mold the training program and set the stage for summative evaluation of immediate program results. Therefore the primary customers of formative evaluation are the instructional designers and developers who are responsible for selecting or developing the instructional performance support system or training package.

Summative Evaluation

Summative evaluation “involves gathering information on adequacy and using this information to make decisions about utilization” (Seels and Richey, 1994, p. 57). Summative evaluation is conducted during or immediately after implementation. There is also a purposeful difference between formative and summative evaluation: “If the purpose of evaluation is to improve . . . then it is formative evaluation. (In contrast, if the purpose is to prove, justify, certify, make a ‘go/no’ decision, or validate . . . then it is summative evaluation.)” (Thiagarajan, 1991, p. 22).

The primary customers are the decision makers who need to approve installation of the instructional performance support system, or in the case of a one-time offering put a final seal of approval on the instructional package. These decision makers may or may not participate in earlier instructional design and development activities. In either case, they need immediate feedback from the first session or the first several sessions: How well did the training meet the stated instructional objectives? How well did it meet expectations of the instructor(s) and participants?

During summative evaluation, “any aspect of the total education or training system can be evaluated: the student, the instructor, instructional strategies, the facilities, even the training organization itself” (Smith and Brandenburg, 1991, p. 35). The designer/developer or evaluator may select from or blend a number of strategies for conducting summative evaluation: cost-benefit analysis, attitude ratings (student, instructor, client, and other stakeholders),

testing (pre-, post-, embedded, and performance tests), surveys, observation, interviews, focus groups, and statistical analysis. The focus is on immediate results; in a situation involving a long-term program, the outputs and outcomes of summative evaluation become inputs for the next step, confirmative evaluation.

Confirmative Evaluation

Confirmative evaluation goes beyond formative and summative evaluation; it moves traditional evaluation a step closer to full-scope evaluation. During confirmative evaluation, the evaluation, training, or HPT practitioner collects, analyzes, and interprets data related to behavior, accomplishment, and results in order to determine “the continuing competence of learners or the continuing effectiveness of instructional materials” (Hellebrandt and Russell, 1993, p. 22) and to verify the continuous quality improvement of education and training programs (Mark and Pines, 1995).

The concept of going beyond formative and summative evaluation is not new. The first reference to confirmative evaluation came in the late 1970s: “The formative-summative description set ought to be expanded to include a third element, *confirmative evaluation*” (Misanchuk, 1978, p. 16). Eight years later, Beer and Bloomer (1986) from Xerox suggested a limited strategy for going beyond the formative and summative distinctions in evaluation by focusing on three levels for each type of evaluation:

1. Level one: evaluate programs while they are still in draft form, focusing on the needs of the learners and the developers
2. Level two: continue to monitor programs after they are fully implemented, focusing on the needs of the learners and the program objectives
3. Level three: assess the transfer of learning to the real world

Geis and Smith (1992, p. 133) report: “The current emphasis is on evaluation as a means of finding out what is working well, why it is working well, and what can be done to improve things.” However, when the quality movement

gained prominence and business thinking raised the bar, educators and trainers began to agree, at least in principle, that “quality control requires continuous evaluation including extending the cycle beyond summative evaluation” (Seels and Richey, 1994, p. 59). Summative evaluation has immediate usefulness, but it does not help planners make decisions for the future. Confirmative evaluation, on the other hand, is future-oriented; it focuses on enduring, long-term effects or results over the life cycle of an instructional or noninstructional performance intervention: “Enduring or long-term effects refer to those changes that can be identified after the passage of time and are directly linked to participation in [education or training]” (Hanson and Siegel, 1995, pp. 27–28).

A Rose by Any Other Name

Since Misanchuk (1978) coined the term, there has been a marked lack of reference to confirmative evaluation. Even so, within the literature related to the design of research there are references to *stability over time*, *repeated measures*, *retention studies*, *recidivism* (tendency to return to a former pattern), and *time series* (S. B. Sawilowsky, personal communication, June 5, 2001). “In education and psychology, terms such as *follow-up studies*, *longitudinal studies*, and *ex-post-facto studies* have reflected the existence of related concepts, as well as the need for such additional evaluations” (Hellebrandt and Russell, 1993, p. 22). In their book on using an ISD model to enhance the role of training in large organizations, Hannum and Hansen (1989) describe two types of evaluation, summative and follow-up: “The second type of evaluation occurs some time after the instruction and is called follow-up evaluation. Its purpose is to evaluate how and if the training is being used by the participants and is used to determine the overall success of the training program. . . . Follow-up data may be collected some months as well as years after participants attend a given training session. . . . Once the data are gathered, they are then analyzed over time to determine overall success and the program is revised as needed” (pp. 36–37).

Rae (1999) uses the term *post-program* evaluation and describes three levels of follow-up evaluation: “Too often evaluation (or what passes for it) does not extend beyond the end of programme validation, but then all that has

been assessed is the satisfaction of the training programme objectives and the immediate objectives of the learners. This goes some way if it has been performed effectively, but it is not complete evaluation and certainly does not lead to an assessment of the value effectiveness of the learning. In order to do this, three further stages are necessary: Post-programme debriefing, medium term evaluation, [and] longer term evaluation” (p. 167).

The quality literature also contains references that imply confirmative evaluation. In addition to the term *continuous improvement*, there are also references to *quality control* and *quality assurance*: “The terms gaining greater acceptance in business training are *quality control* for input and process functions and *quality assurance* for product or output functions” (Brandenburg, 1989, pp. 85–86). As Seels and Richey (1994) state, “The quality improvement movement will affect the evaluation domain. Quality control requires continuous evaluation including extending the cycle beyond summative evaluation” (p. 59).

Two other terms related to confirmative evaluation are *outcome evaluation* and *impact evaluation*. According to Schallock (1995), outcome-based evaluation is “a type of program evaluation that used valued and objective person-referenced outcomes to analyze a program’s effectiveness, impact, or cost-benefit” (p. 5). Further, impact evaluation looks at negative or positive program-based changes in performance and focuses on “whether the program made a difference compared to either no program or an alternate program” (p. 6). Although outcome and impact evaluation are not synonymous with confirmative evaluation, confirmative evaluation does contain elements of both outcome and impact evaluation.

Even level four of Kirkpatrick’s four levels of evaluation (1959, 1994) is confirmative evaluation by another name. Level four measures the results of training in terms of change in participant behavior and “tangible results that more than pay for the cost of training” (1994, p. 69).

Meta Evaluation

Formative, summative, and confirmative evaluation are all fodder for meta evaluation. Meta evaluation “. . . is a quality control process that is applied to the processes, products and results of formative, summative, and confirmative

evaluation” (Van Tiem, Moseley, and Dessinger, 2000, p. 181). It is all about evaluating the evaluation. The evaluator literally zooms in on how the evaluation was conducted. The purpose of meta evaluation is to validate the evaluation inputs, process, outputs, and outcomes. It serves as a learning process for the evaluator and makes the evaluators accountable: “Evaluators will be more likely to see their studies effectively utilized when they demonstrate that their work can stand the test of careful analysis and that they themselves are open to growth through constructive criticism” (Posavac and Carey, 1989, pp. 281–282).

There are two types of meta evaluation: type one and type two. Table 1.1 describes the two types of meta evaluation in terms of timing and purpose.

Table 1.1. Meta Evaluation: Type, Timing, and Purpose.

Type	Timing	Purpose
Type one (formative)	Conducted during formative, summative, and confirmative evaluation	Guides the evaluator through the planning, design, and implementation of all three stages of evaluation
Type two (summative)	Conducted after the formative, summative, and confirmative evaluations are completed	Provides feedback on the reliability and validity of the evaluation processes, products, and results

Source: Van Tiem, D. M., Moseley, J. L., and Dessinger, J. C. *Fundamentals of Performance Technology: A Guide to Improving People, Process, and Performance*. © 2000 International Society for Performance Improvement. p. 181. Reprinted with permission.

Type one meta evaluation is conducted concurrently with the evaluation process. It is literally a formative evaluation of evaluation. Type two meta evaluation is the more common approach. It is conducted after formative, summative, and at least one cycle of confirmative evaluation is completed. Stufflebeam (1978) and The Joint Committee on Standards for Educational Evaluation (1994) offer an extensive set of utility, feasibility, propriety, and

accuracy standards and guidelines for conducting a type two meta evaluation in education or training settings.

Comparing the Four Types of Evaluation

Table 1.2 illustrates a comparison of the four types of evaluation. The constructs used for comparison are timing, purpose, and customers.

Table 1.2. Evaluation Types: Timing, Purpose, and Customers.

Type	Timing	Purpose	Customers
Formative	During design, development, and pilot or field testing	Improve analysis, design, development processes, and outputs	Primary: design team (designers, developers, instructors, subject matter experts, etc.) Secondary: decision makers and customers
Summative	During or immediately after full implementation	Assess immediate results (outputs and outcomes)	Primary: decision makers and customers Secondary: design team (designers, developers, instructors, subject matter experts, etc.)
Confirmative	3-12 months after full implementation	Assess effectiveness, efficiency, impact, and value over time	Primary: decision makers and users Secondary: design team (designers, developers, instructors, subject matter experts, etc.)
Meta	Type one: concurrent with development and implementation Type two: after development and implementation	Validate evaluation process, products, outputs	Primary: evaluators Secondary: decision makers and users

Source: Dessinger (2002).

Timing

Like summative evaluation, confirmative evaluation takes place after development and implementation. Hellebrandt and Russell (1993) state that confirmative evaluation should occur six months to one year after development and initial implementation, depending on the criticality, complexity, and frequency of the learning or performance. Carr (1992, p. 151) is even more aggressive, suggesting that confirmative evaluation of ongoing training programs “should begin the day the first training ends.”

If we consider confirmative evaluation in terms of *assessing impact*, Rossi, Freeman, and Lipsey (1999) insist that “interventions should be evaluated for impact only when they have been in place long enough to have ironed out implementation problems” (p. 238). Implementation problems may include failure to deliver critical elements to appropriate targets, lack of measurable outcomes, or lack of summative or formative evaluation data. For example, during the implementation of a recent leadership training program, individual instructors selected modules that they felt best suited the audience for a particular session; they did not present the entire program. Some instructors also decided not to conduct summative evaluation at the end of their sessions because the sessions ran over the allotted time and the participants were eager to leave.

Performance Support Tool (PST) 1.1 is a decision matrix that helps the practitioner decide when to implement a confirmative evaluation on the basis of the criticality, complexity, and frequency of the training program’s intended learning or performance outcomes.

PST 1.1. When to Conduct a Confirmative Evaluation.

Purpose: To help you decide when to conduct a confirmative evaluation on the basis of the criticality, complexity, and frequency of the learning or performance.

Directions: Ask all the stakeholders to rate the training program according to the criteria in the first column. Then decide when to confirm the outcomes of the training program.

Learning/Performance Factor	Rating	Confirm Every . . .
Criticality: How critical is the learning/performance to the success of the organization?	<input type="checkbox"/> High	<input type="checkbox"/> 3–6 months
	<input type="checkbox"/> Medium	<input type="checkbox"/> 6–12 months
	<input type="checkbox"/> Low	<input type="checkbox"/> 12 months
Complexity: How complex is the learning/performance?	<input type="checkbox"/> High	<input type="checkbox"/> 3–6 months
	<input type="checkbox"/> Medium	<input type="checkbox"/> 6–12 months
	<input type="checkbox"/> Low	<input type="checkbox"/> 12 months
Frequency: How often is the learning/performance required?	<input type="checkbox"/> Regularly	<input type="checkbox"/> 6–12 months
	<input type="checkbox"/> Monthly	<input type="checkbox"/> 6–12 months
	<input type="checkbox"/> Annually	<input type="checkbox"/> 12 months
	<input type="checkbox"/> One time only	<input type="checkbox"/> Do not confirm
Frequency: How often is the learning/performance implemented?	<input type="checkbox"/> Regularly	<input type="checkbox"/> 6–12 months
	<input type="checkbox"/> Monthly	<input type="checkbox"/> 6–12 months
	<input type="checkbox"/> Annually	<input type="checkbox"/> 12 months
	<input type="checkbox"/> One time only	<input type="checkbox"/> Do not confirm

Source: Van Tiem, D. M., Moseley, J. L., and Dessinger, J. C. *Fundamentals of Performance Technology: A Guide to Improving People, Process, and Performance*. Copyright © 2000 International Society for Performance Improvement. p. 180. Used with permission.

Suggested start-up targets range from three months to one year after the implementation of the training program. If the learning or performance outcomes of the training program are highly critical to the success of the organization or are very complex, the program should undergo confirmative evaluation between three and six months after implementation. On the other hand, an evaluator may wait a year to conduct confirmative evaluation of a training program whose learning or performance outcomes are rated low in criticality or complexity. There is no need to conduct a confirmative evaluation if the learning or performance outcomes take place only one time or infrequently—for example, a training program on how to develop an organization's five-year strategic plan.

Purpose

Formative and summative evaluation each zoom in on needs, processes, products, reactions, and accomplishments. The purpose of formative evaluation is to validate the needs analysis and the training program design process and outputs; the purpose of summative evaluation is to assess the participant's accomplishments during and immediately after training. Confirmative evaluation zooms out to take a long-term view of the effectiveness, efficiency, impact, and value of a training program. The purpose of confirmative evaluation is twofold: (1) identify, explain, and confirm or justify the continuing value of training and learning over time; and (2) help decision makers manage the instructional performance support system and the learner over time (Van Tiem, Moseley, and Dessinger, 2000).

Customers

The primary customers for formative evaluation are the program designers and developers; the primary customers for summative and confirmative evaluation could include any or all of the internal and external stakeholders, that is, anyone who has a vested interest (expressed as need or expectation) in the process, outputs, and outcomes of the training program. Shrock and Geis (1999) support the concept that the customer base for summative evalua-

tion may be broader because traditionally “much of the feedback used to make revisions to an intervention is collected after an intervention is implemented” (p. 192).

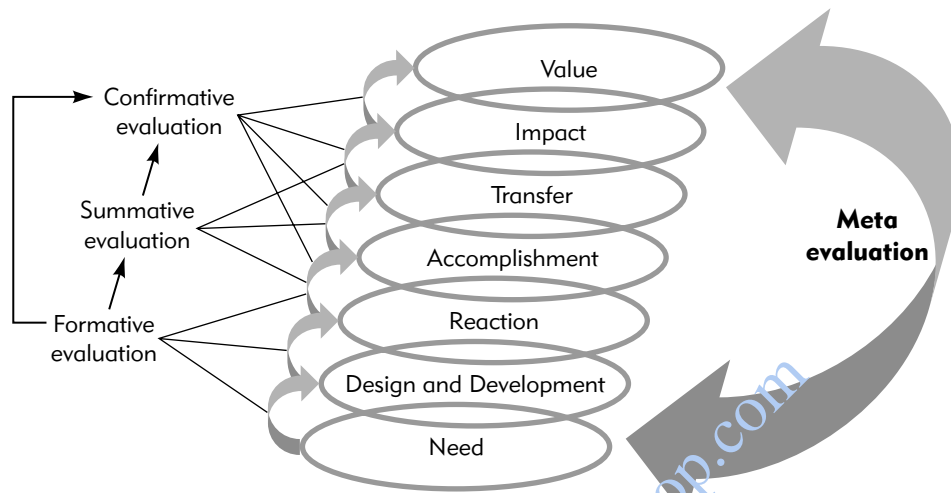
The primary customers of confirmative evaluation are long-term decision makers. Executives, managers, consultants, and others may use the outputs and outcomes of confirmative evaluation for strategic planning. Other decision makers use the results of confirmative evaluation to determine whether to maintain, improve, discard, or replace the training program or noninstructional performance intervention. In addition, decision makers involved with certification processes have a special stake in confirmative evaluation outcomes. For example, “The notion of confirmative evaluation is significant in the health professions in terms of assuring that learners maintain their clinical knowledge and skills. . . . Confirmative evaluation in nursing is significant, particularly in the clinical setting, to assure that learners maintain their competencies and to identify where additional review and practice are needed” (Oermann and Gaberson, 1998, p. 5).

Evaluation: Full-Scope Model

Going beyond formative and summative evaluation “. . . challenges us to jettison linear models and integrate evaluative processes throughout every phase of ISD” (Moseley and Solomon, 1997, p. 12), not just add evaluation at the end, as implied by the traditional ADDIE (analyze, design, develop, implement, evaluate) model. The Dessinger-Moseley 360° Evaluation Model (Moseley and Dessinger, 1998) illustrates the integration of evaluation throughout the ISD process (ADDI/E) and presents six foci for integrating evaluation activities: need, design and development, reaction, accomplishment, transfer, and impact. The model was influenced by the work of Seels and Richey (1994); Kirkpatrick (1994); Kaufman, Keller, and Watkins (1996); and Brinkerhoff (1987).

Figure 1.1, the Dessinger-Moseley Full-Scope Evaluation Model, expands the earlier 360° model by adding *value* to the foci and by introducing the concept of full-scope evaluation.

Figure 1.1. Dessinger-Moseley Full-Scope Evaluation Model.



The spiraling concentric circles of the Full-Scope Evaluation Model represent the “proactive and iterative nature of evaluation” (Moseley and Dessinger, 1998, p. 247). Within the circles are the focal points for each type of evaluation:

- Need for training
- Instructional design and development processes and products
- Reaction during and after training
- Accomplishment during training
- On-the-job transfer of new knowledge, skills, or attitudes
- Impact of negative or positive training results on the individual, business group, organization, or global community
- Value to the individual, business group, organization, and global community

The center of the model illustrates the foci for formative, summative, and confirmative evaluation:

- Formative evaluation focuses on need, design and development, and pre-implementation reaction and accomplishment
- Summative evaluation focuses on the immediate results of program implementation: reaction; accomplishment; and the self-reported expectation that new knowledge, skills, and attitudes will transfer to the job and affect workplace performance
- Confirmative evaluation focuses on the program's continuing impact and value, as well as long-term transfer, accomplishment, and reaction

The left side of the model emphasizes that input from formative evaluation flows into summative and confirmative evaluation; input from summative evaluation flows into confirmative evaluation, and each type of evaluation has its own set of evaluation foci. On the right side of the model, the process of meta evaluation focuses on all the evaluation types—their inputs, outputs, outcomes, and foci.

Challenges to Full-Scope Evaluation

Full-scope evaluation is not without challenges. Daring to go beyond the traditional formative and summative framework is the first challenge. Added to this are demands created by the new organization of the twenty-first century, context factors specific to an organization, and the challenge of overcoming the “success syndrome.” All these challenges are interrelated and call for strong action from the evaluation, training, or HPT practitioner.

Daring to Go Beyond Formative and Summative Evaluation

Full-scope evaluation should be the norm rather than the exception. More and more, training is considered an integral part of strategic planning. As investments in training increase, there is a corresponding increase in the

“expectation that workplace improvement practitioners rigorously measure the outcomes that these investments produce, and in so doing, generate the insight and understanding necessary to continuously improve those outcomes” (Bassi and Ahlstrand, 2000, p. 1). However, even more enlightened organizations still view evaluation as a costly add-on rather than a value-added activity. They think of full-scope evaluation in terms of how much the additional time, money, and other resources will cost. The practitioner may even find that “either there is no positive consequence for human resources or training in demonstrating business results, or there are actually disincentives” (Binder, 2002a, p. 8). It’s hard to dare to think outside the box when the organization does not support, or even punishes, such thinking.

Adapting to the New Organization

Despite current lack of support for full-scope evaluation, futurists see a new organization emerging in the twenty-first century. Reed (2002, pp. 24–25) suggests that this new organization requires a new way of looking at training: “Training’s definition should be the provision of learning opportunities for successful performance improvement . . . training must also be seen as a process that is continuous and learner centric; one that focuses on the pull or output side.” The focus on results challenges organizations and training departments to take a new look at evaluation in general and confirmative evaluation in particular. The practitioner needs new knowledge and skill to function as a change agent and a cheerleader for full-scope evaluation.

Adjusting for Context Factors

Context factors also create a challenge. Context factors include organizational culture, climate, and environment and are often codified in the organization’s mission, values, and goals. To implement full-scope evaluation, the total organization must:

- Recognize the long-term value of education, training, and learning
- Actively support the concepts of accountability and continuous improvement

- Recognize the value of full-scope evaluation of education, training, and learning
- Commit to full-scope evaluation
- Actively support full-scope evaluation

In turn, the evaluation, training, or HPT practitioner, and stakeholders who are involved in the planning and implementing of full-scope evaluation must:

- Know the mission, goals, and values of the organization
- Agree with the mission, goals, and values of the organization
- Value and buy into the mission, goals, and values of the organization
- Support the organization by aligning evaluation outcomes with the mission, goals, values, and culture of the organization

“Know your organization” is the caveat and challenge that can guide internal and external evaluation, training, and HPT practitioners to the successful implementation of full-scope evaluation.

Overcoming the “Success Syndrome”

One final challenge is the success syndrome: the tendency for individuals and organizations to use positive outputs from summative evaluation to trigger a decision not to proceed with full-scope evaluation. Consistently high positive participant reaction forms or consistently low pretest and high posttest scores are the major culprits.

In one example, decision makers from a division of a major automotive company approved a full-scope evaluation plan for a new, long-term, basic-skills training program. However, the early summative evaluation results at level one (did they like it?) and level two (did they learn?) were so positive that the decision makers decided to save money by discontinuing *all* types and levels of evaluation. The training program continued unchanged for three more years, and then the participant materials were distributed over the company intranet as part of new-employee orientation. Finally, the online program was absorbed into a corporate university curriculum, where it died a natural death—no one signed up for the course.

