CHAPTER 13

Morrison, G., Ross, S., Kalman, H. and Kemp, J. (2011). Designing Effective Instruction (pp. 342-377). New York: Wiley.

Using Evaluation to Enhance Programs: Conducting Formative and Summative Evaluations

GETTING STARTED

You receive a contract from a software firm to evaluate a series of computer-based instructional units for teaching high school geometry. The units are still in draft stage, with the expectation that they can be revised based on the results of the evaluation. Using the eight-step model (see Figure 13-1), you determine the purposes of the evaluation, conduct an analysis of the audience, identify issues (questions and objectives), determine available resources, identify the evidence that will be needed, specify and implement data-gathering techniques, conduct the data analysis, and write and present reports of the findings. To address the main interests of the stakeholders (i.e., the software developers), you decide to employ small-group trials in which students work through the materials, give "think-aloud" reactions and other feedback, and take unit achievement tests. You also use two SMEs, three high school geometry teachers, and one CBI design expert to give impressions of the accuracy and quality of the content and user interface. Based on qualitative and quantitative analyses of the multiple data sources, you prepare a report detailing results and making specific recommendations for improving various parts of the units. One of the managers in the software firm criticizes your evaluation for "failing to prove whether the units actually increase learning." She further wants to know why you didn't use a control group. How would you respond to the manager's concern?

QUESTIONS TO CONSIDER

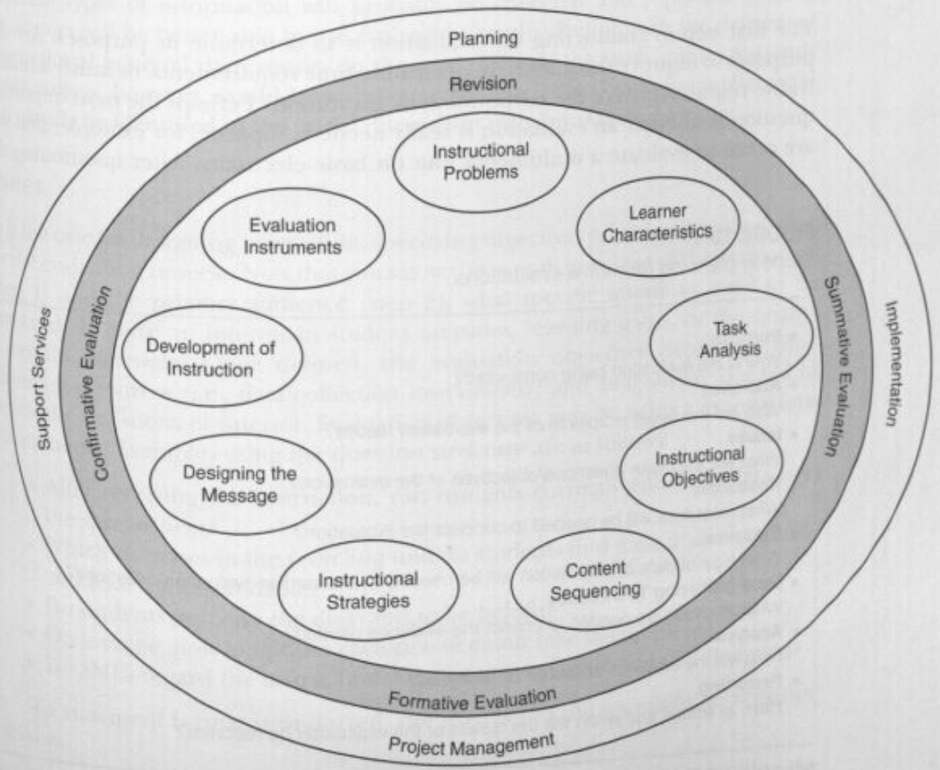
"At what stage of instructional development does formative evaluation begin?"

"How can formative evaluation results be used to improve instruction?"

"How can an instructional designer provide evidence that systematic instructional planning does pay off?"

"What is the actual cost of an instructional program?"

"How can a training program be valuable if it doesn't directly produce income for the company?"



In Chapter 11, we differentiated between the three broad categories of evaluation—formative, summative, and confirmative—used by instructional designers. Now that you have a better understanding of the functions of evaluation as well as instrument construction, we turn to the procedures for conducting each type of evaluation.

A BASIC MODEL FOR FORMATIVE EVALUATION

Formative evaluations are used to provide feedback to designers for making course improvements (Sterbinsky & Ross, 2005). These evaluations take place as instruction is "forming," and thus they precede the development of the final version of the instructional unit or course. As is true for the design of instructional material, formative evaluations must be carefully planned to be effective. To help structure the evaluation planning, Gooler (1980) suggests the eight-step approach summarized in Figure 13-1. Each of the steps is examined next.

Purposes

The first step in conducting the evaluation is to determine its purposes: Are the purposes to improve the materials, determine time requirements, or satisfy administrative requirements of the corporation or institution? Perhaps the most important question is whether an evaluation is really needed. Suppose, for example, that you are asked to evaluate a multimedia unit on basic electricity. After questioning the

FIGURE 13-1
Steps in planning formative evaluations

- **Purpose**
  Why is the evaluation being conducted?
- **Audience**
  Who are the target recipients of the evaluation results?
- **Issues**
  What are the major questions/objectives of the evaluation?
- **Resources**
  What resources will be needed to conduct the evaluation?
- **Evidence**
  What type of data or information will be needed to answer the evaluation questions?
- **Data-gathering Techniques**
  What methods are needed to collect the evidence needed?
- **Analysis**
  How will the evidence collected be analyzed?
- **Reporting**
  How, to whom, and when will the results of the evaluation be reported?

Adapted from D. D. Gooler (1980), "Formative Evaluation Strategies for Major Instructional Development Projects," *Journal of Instructional Development, 3,* 7–11

course administrators, you discover that the unit cannot be changed without the software being rewritten. You further determine that there are no available funds to pay for new programming. Although there still may be good reasons for doing a formative evaluation, you certainly seem justified in questioning its purpose at the front end. That is, if the evaluation results cannot be used to make changes, is this really the best time for an evaluation?

The purposes of the evaluation are usually defined through consultation between the evaluator and the stakeholders of the course or program. As the name implies, stakeholders are individuals who have a "stake," or vested interest, in the instruction. They might include company or school administrators, course vendors, training professionals, and/or teachers or trainers.

Audience

An additional part of the initial planning is to determine the intended audience(s) for the evaluation results. Will they be managers, teachers, course developers, or a combination of several groups? Depending on who the primary audience is, different types of information will probably be collected and reported. Clearly, instructors will be better able to use evaluation results dealing with the delivery of instructional material than results on the readability of the study guide; the study guide author, however, would have the opposite need. The key target audience(s) will usually be identified in the initial discussions with key stakeholders.

Issues

As is the case for designing instruction, specifying objectives provides the foundation for the evaluation process. Now that you know the overall purposes of the evaluation (step 1) and the primary audience (step 2), what specific questions need to be answered? Is there an interest in student attitudes, learning gains, or the quality of certain materials? Once defined, the evaluation objectives determine what information sources (i.e., data collection instruments) and analyses are needed to answer the questions of interest. Evaluation objectives may be written as questions or statements. Examples using the question structure are as follows:

- After receiving the instruction, can students correctly enter the data into the spreadsheet?
- Which exercises in the recycling unit do students find most understandable and least understandable?
- Do students perceive the diagrams to be helpful?
- On average, how long does each practice unit take to complete?
- Do SMEs regard the instructional material as accurate and well designed?

If a statement format is preferred, the first two examples might be rewritten as follows:

- To determine students' accuracy, following the instructions, in entering the data into the spreadsheet

- To identify the exercises in the recycling unit that students find most understandable and least understandable

Resources

Given the evaluation objectives (step 3), what resources are needed to address each? In the previous example, objective 1 implies the need for students to be tested, the need for computers and spreadsheets for the students to use during testing, and the need for keyboarding test(s) for measuring the degree of change from pre- to postinstruction. Objective 2 also involves gathering data from students, but this time using a survey or interview to determine the exercises they most and least prefer. The resources needed therefore differ from objective to objective. Common types of resources include these:

- Trainees/students
- SMEs
- Instructors
- Data collection and analysis instruments
- Copies of materials
- Physical facilities and equipment

Evidence

In conjunction with identifying resources, careful consideration must be given to the types of evidence that will be acceptable for addressing the evaluation objectives. For objective 1 in our example, we obviously want to obtain keyboarding scores, but will scores from, say, five students on one test suffice, or will additional students and/or multiple testings be required? For objective 2, thought must be given to the type of student reporting (in reacting to the exercises) that will be most valid and informative. We might be skeptical, for example, about the validity of impressions conveyed in an interview immediately following a difficult final exam. In deciding what will constitute acceptable evidence, the evaluator and the stakeholders may want to consider these points:

- Sample size
- Objectivity of the information sources
- Realism of the testing context
- Degree of control in the testing context
- Need for formal statistical reporting
- Reliability/validity of SME reviews

Data-Gathering Techniques

This step involves making final decisions about the instrumentation and data collection methods to be employed. Two key, and often opposing, factors need to be weighed: precise measurement versus feasible or practical measurement. For example, in planning a formative evaluation of a carpentry training program, the designer initially elects to rate actual projects that students complete on the job. But once the practical problems of identifying and validly assessing actual projects are considered, it is decided instead to employ a controlled assessment, specifically, a cabinet door of a specified design and size completed at the training site. Similarly, to reduce time and cost, planned interviews with 30 students may be reduced to include only 10 students. Or perhaps the designer may substitute an interview for a questionnaire when he or she considers that the former provides opportunities to probe for more in-depth explanations but is more time consuming.

The data-gathering techniques available were discussed in the two previous chapters. They include performance tests, written tests, observations, ratings, questionnaires, interviews, portfolios, and exhibitions. At this stage, the designer should consider the advantages and limitations of each type for addressing the evaluation objectives as well as the resources necessary for each. In Chapters 11 and 12, we recommended the use of multiple measures to increase validity (through the triangulation of findings across measures) and the provision of as much relevant information as feasible. Cognitive and constructivist paradigms have increased awareness about obtaining data regarding learning processes as well as products (Kay & Knaack, 2009; Ross & Morrison, 1995). By knowing about processes—how instructional material is used—designers are in a better position to interpret products (i.e., outcomes) and thereby improve the material.

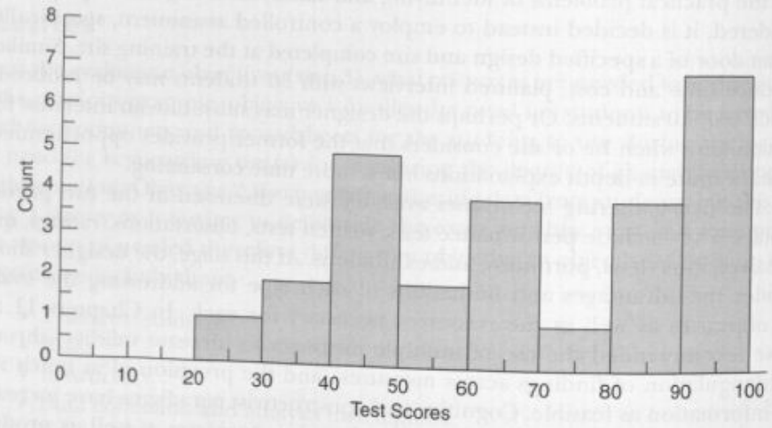
Analysis

Once the data are collected, the next step is analyzing the results. If the term *analysis* conjures up thoughts of complex statistics and formulas, reflect again on the main purpose of formative evaluation. It is to provide usable information for designers to improve instruction. Although the need for complex analyses should not be ruled out, the questions of interest are often best addressed by straightforward and fairly simple descriptive analyses. These types of analyses generally tell us how students performed on or reacted to a particular lesson. Typical analysis procedures include the following. Note that the first three are quantitative (i.e., involve numerical indices), and the fourth is qualitative (i.e., involves impressions); both types of analyses are discussed later.

- Frequency distributions
- Frequency graphs or histograms
- Descriptive statistics, such as percentages, means, and medians
- Listing of actual comments made by respondents

Figure 13-2 shows a frequency distribution and descriptive statistics for a class on a unit achievement test (maximum score = 100%). Note that the "bars" on the graph represent the number (i.e., frequency) of scores obtained within 10-point intervals (20–29, 30–39, etc.). From the distribution, it is clear that performances were quite spread out (range = 75) and also somewhat low for many students

FIGURE 13-2
Histogram of test score



Mean:	Std. Dev:	Std. Error:
68.077	22.765	4.465
Minimum:	**Maximum:**	**Range:**
23	98	75

(mean = 68.1). Although it is possible that these are desirable performances for this unit (i.e., it may be very difficult material), chances are, the designer will not be satisfied and will conclude that the instruction needs to be revised.

In many cases, an introductory-level knowledge of statistics will suffice for completing the data analysis. If more complex analyses are required, the designer can always seek assistance from a statistical consultant. Although traditional views of data analyses denote graphs, tables, numbers, and probability values, today's educational evaluators may find that they depend just as much (or more) on qualitative analyses. These types of analyses involve categorizing, interpreting, and, in general, "making sense" out of subjective data such as observations of instruction or student learning, interview responses, and open-ended survey responses. Qualitative researchers who are collecting data for scientific study and intended wide-scale dissemination of results will want to invest considerable effort into ensuring high reliability in their methods, inferences, and conclusions. Consequently, the thorough and systematic procedures for qualitative data analysis recommended by such authors as Miles and Huberman (1994) should be followed. The qualitative evaluator of instruction will also want to ensure valid conclusions, but may, because of time constraints and

project needs, use a more practical, less rigorous approach. Key strategies should include the following:

- Reviewing notes and transcripts to extract major categories or themes (e.g., "The most common concerns about the training unit were inadequate length, poor readability of materials, and lack of relevance of the unit on reflective listening.")
- Providing a sense of the saliency or importance of the themes (e.g., "The inadequate length of the training was a critical weakness, as conveyed by the majority of trainees; only several, however, were concerned about the readability of materials.")
- Conducting "member checking" by having respondents review and validate, where feasible, your description and interpretation of what they said
- Providing tables that present typical responses or a complete listing so that stakeholders in the instructional design might review them
- Having more than one person participate in the data analysis so that reliability can be checked and increased by consensual agreement

Reporting

The evaluation effort will generally be of little value unless the results are disseminated to individuals involved in the instructional unit or course (e.g., instructors, administrators, or designers). The most common means is the evaluation report. There is no single, standard reporting format; in fact, the best strategy is to adapt the report to the primary target audience with regard to content and style of writing. A "typical" report, however, is likely to include most or all of the following sections:

I. Executive summary (abstract)
II. Purposes of evaluation
  A. Evaluation objectives
  B. Description of target course/unit
III. Methodology
  A. Participants
  B. Instruments
IV. Results
  A. Analyses
  B. Findings
V. Conclusions and recommendations

A second common dissemination approach is oral reporting. Depending on the context, such reports may be formal presentations, group meetings, or one-to-one discussions. Keep in mind that, whatever the form of reporting, the overall goal of formative evaluation is to recommend and make changes (as suggested

by the results) to improve instruction. Formative evaluation results are unique to the particular project and thus will have limited generalizability to projects outside the same educational context or curriculum.

TYPES OF FORMATIVE EVALUATION

As just described, the initial planning clarifies the purpose(s) of the evaluation and the target audience(s) for receiving the evaluation results (e.g., the design team, funders, potential clients, etc.). These decisions then dictate the evaluation approach that is most appropriate for the particular project. According to Flagg (1990), the most commonly used evaluation approaches can be classified into four categories: connoisseur-based, decision-oriented, objectives-based, and public-relations-inspired studies.

Connoisseur-Based Studies

A connoisseur-based study employs SMEs and other appropriate consultants (e.g., media and design experts) to examine the instruction and give opinions regarding its accuracy and effectiveness. An important part of the expert's report is any recommendation for revising the instruction where improvements are needed. Expert review can be useful at all stages of the design process, from initial drafts of instructional material to the completed versions.

In using expert review, the experts are assumed to be competent and interested in the evaluation task. Important tasks for the designer are to determine the number and types of experts needed, the particular individuals who will fill these slots, and the best time(s) in the evaluation process to involve each. Expert review can provide valuable information for refining instruction. However, an important limitation of connoisseur-based studies is that so much depends on the biases and experiences of the selected experts. Remember, experts in a subject area are not necessarily knowledgeable about instruction and learning. Therefore, it is not advisable to blindly follow a recommendation that you question without seeking other opinions or data sources. In educational research, this type of verification process is called triangulation, the procedure of cross-validating a finding by using multiple information sources.

Consider this example: After reviewing a CBI lesson, a media expert suggests that children will not know which keys to press in selecting different branching options. A field test of the program reveals, however, that this problem does not occur—the children actually make very few errors. The implication: Keep the expert's concern in mind, but do not make changes in the screen design without further evidence.

It is also important that the expert is willing and able to provide accurate and objective feedback. In this regard, consider the situation described in the Expert's Edge section.

Expert's Edge

Never Let a Fox Guard the Henhouse Even If He Claims to Be a Vegetarian

The workshop had been designed by a well-known training organization and followed a systematic product realization process. The process included several "gates," each of which specified a formative evaluation requirement. Early gates required reviews by SMEs, clients, and instructional designers, whereas the penultimate gate required a formal field trial. By all accounts, the workshop passed each gate with flying colors. The problem was, however, that participants' feedback on the workshop was not as positive as the client had hoped, and the work products participants created as a result of the workshop lacked the connection to business results the client had intended. Further investigation revealed that the formative evaluation process, although it had been followed to some extent, had been co-opted. In the interest of meeting the target dates in the project plan, "friendly" SMEs and client representatives had been selected for the reviews, and the instructional design review had been done by the project manager. Furthermore, the field trial was actually a run-through of workshop materials not before representative members of the target audience but before a small group of colleagues of the project manager. Conversations with those involved in the development of the workshop showed no malicious intent; as a matter of fact, there was no real conscious awareness that the process had been subverted. Eventually, the workshop was redesigned following a more stringent process. But considerable damage had been done, and because of the rework required, the financial cost was significant. Reputations were sullied, client trust was weakened, work was reassigned, and jobs were lost.

Formative evaluation can take many forms, and it can follow varying degrees of formality. Despite immediate pressures to the contrary, formative evaluation need not be incompatible with speed or staying within the budget. But formative evaluation is a *sine qua non* of our profession, and ultimately we bear the responsibility to our clients and to ourselves to conduct these activities with rigor and discipline. Here are some actions you might take in this regard:

- Build the time and cost of formative evaluation into your project plans. Make sure your clients and management understand in their own terms the value of formative evaluation.
- Design formative evaluations so that subversion is unlikely—use outside evaluators and employ audits or evaluation panels, for example.
- Develop and strive to maintain a set of personal standards by which your work can be known.
- Create an atmosphere where the truth can be told.
- Walk away from projects where your standards cannot be upheld.

Robert Bowman is an instructional development and evaluation manager in the Lucent Technologies Learning and Performance Center. His current assignment focuses on assessing the impact of leadership development programs and converting leadership development programs to technology-based systems. His work has included a broad range of needs assessment, instructional development, and evaluation assignments.

Decision-Oriented Studies

Decision-oriented studies are designed to provide information related to particular questions about the instruction. Example concerns might include these:

- Does the lesson require too much time to complete?
- Is the amount of practice examples adequate?
- Are prerequisite math skills required?
- Should a student study guide be developed?

Given these questions, the evaluator then designs specific measures and procedures to address each. Decision-making evaluations are naturally most valuable when program changes are still feasible and economical. Good communication between the evaluator and the program stakeholders is essential to define the relevant questions around which the decision making will be based. The limitation of decision-making studies is that their results are usually descriptive, not prescriptive. That is, they tell us how the instruction is working but not what to do to make it better.

Objectives-Based Studies

A third category of formative evaluation approaches involves investigating how well the instructional program is achieving its objectives. The basic methodology therefore resembles that employed in summative and confirmative evaluations (discussed later in this chapter) by assessing the amount of progress students have realized from completing the instruction. Summative and confirmative evaluations, however, use such information to judge the effectiveness of the program immediately (summative) and over time (confirmative). Objectives-based studies are an appropriate choice for evaluating a project such as a distance education course to determine whether it is effective. Formative evaluation uses them as a basis for improving the instruction where outcomes fall short of goals.

Objectives-based studies frequently employ pretest–posttest designs that measure gains on measures of achievement and attitude. Their main limitation is the same one noted earlier for decision-making studies: The findings by themselves provide limited direction for making improvements. For that reason, it may make sense to combine expert opinion (i.e., connoisseur-based study) with the objectives-based results.

Public Relations-Inspired Studies

By making evaluation results known to targeted individuals, public-relations-inspired studies are used to solicit financial support or backing for a project. An example comes from one of the authors' recent experiences in trying to obtain funding from a private foundation to support an elementary school reading program. By conducting a formative evaluation of a pilot version of the program, he was able to present preliminary data that convinced the foundation representatives of the program's potential. The result was a foundation grant to expand both the program and the evaluation the next year.

These four categories should be viewed as complementary rather than exclusive. Most studies will employ combinations of two or more of these orientations, depending on the evaluation objectives. For example, in evaluating a new mathematics program for lower-achieving children, the designers might (1) employ experts to review the instructional materials (category 1), (2) administer surveys and use observation techniques to answer questions about the implementation (category 2), (3) administer achievement pretests and posttests to assess the level of reading improvement demonstrated by student participants (category 3), and (4) publicize the latter results to obtain additional funding for the program (category 4).

Constructivist-Oriented Process

Are traditional formative evaluation approaches too limited and behaviorally oriented to assess contemporary Web-based instructional tools? Kay and Knaack (2009) raise this concern for evaluating what they generically label "leaning objects." Such support tools, which encompass help-systems prompting, hierarchical navigation aids, program-generated questions, graphics, and animations, are not designed to simply increase the amount of content remembered (Friesen & Anderson, 2004; Krauss & Ally, 2005; Nurmi & Jaakkola, 2006) but to engage students in higher-order learning and problem solving. The formative (and summative) evaluation model they propose uses a combination of the connoisseur-based and decision-oriented approaches to obtain *valid* triangulated data addressing the three factors of *learning construct* (e.g., interactivity, feedback, graphics, etc.), *quality* of the instruction (e.g., help supports, instructions, organization, etc.), and *engagement* (e.g., interest, motivational, enjoyment, etc.). Data sources are students participating in field trials who complete a survey, make open-ended comments, and complete pretests and posttests on the content of the instruction. Teachers, in turn, implement the field trial while answering questions on the degree of student learning, instructional quality, and student engagement.

This multifaceted model differs from traditional formative evaluation designs by examining broader, constructivist properties of the instruction. Depending on preferred theoretical and operational preferences of the designer, this approach may be considered as a useful paradigm for the increasing design work being performed to accommodate open-ended and Web-based learning environments (e.g., Clarebout & Elen, 2008; Saab, Gijlers, Van Joolingen, & van Hout-Wolters, 2008).

STAGES OF FORMATIVE EVALUATION

At this point, we have presented general approaches and a specific procedural model (see Figure 13-1) for conducting formative evaluations. A remaining question may be how formative evaluation methods might vary at different stages of the design process. Dick and Carey (1991) addressed this issue in their three-stage model (see Table 13-1).

TABLE 13-1
Stages of formative evaluation

Stage	Instruction Phase	Purpose	Learners	Main Measures
One-to-one trials	Development	Try-out impressions	Individuals	Observation, survey, interview
Small-group trials	Preliminary/draft version	Identify strengths/weaknesses	Small groups (8–20)	Observation, attitudes, performance
Field trials	Completed	Assess actual implementation	Regular classes	Performance, attitudes

The first stage, occurring toward the beginning of the process and usually repeated several times, consists of one-to-one trials, or developmental testing (Thiagarjan, Semmel, & Semmel, 1974) in which the designer "tries out" the instruction with individual learners (Brenneman, 1989). The goal is to obtain descriptive information pertaining to the clarity, impact, and feasibility of initial versions of the instruction.

The second stage consists of small-group trials in which a more developed version of the instruction is used with a group of between 8 and 25 individuals. Through observational, attitudinal, and performance data, the evaluator attempts to identify strengths and weaknesses in the instruction before it is put into a "final" form.

The third stage is the field trial, which examines the use of the instruction with a full-sized learner group under realistic conditions. Based on the results from various outcome measures, the instructor would make final revisions and deliver the completed instruction to actual classes. But at this point, the need for evaluation is not over. Summative and confirmative evaluations are then required to determine whether the instructional program is achieving its goals.

SUMMATIVE EVALUATION: DETERMINING PROGRAM OUTCOMES

Too often an instructional designer or an instructor may intuitively be convinced that what is being accomplished is worthwhile and successful. It is often assumed by persons in education and training that the merits of a program are obvious to other persons in the institution or organization. Unfortunately, rarely is either of these conclusions true.

A summative evaluation permits a designer or instructor to reach unbiased, objective answers to evaluation questions concerning expected program outcomes and then to decide whether the program is achieving those outcomes. With this evidence, the designer's intuition can be supported or rejected, and he or she has the facts for correctly informing others about the program results. The following important issues can be examined through summative evaluation procedures:

- Effectiveness of learner or trainee learning
- Efficiency of learner or trainee learning

- Cost of program development and continuing expenses in relation to effectiveness and efficiency
- Attitudes and reactions to the program by learners, faculty, and staff
- Long-term benefits of the instructional program

In this section, we examine methods for gathering data that can lead to a conclusion for each of the five issues stated earlier. Attention to these matters may be essential in proving the value of a new instructional program and then ensuring its continued support.

A summative evaluation of a course or program is more than a one-time activity. Immediately after each course or training program is concluded, the instructor should utilize some or all of the assessment methods described subsequently. By accumulating summative data, continuing positive trends in a program can be tracked over time, or deficiencies can be noted as they show up and corrections made immediately (Ross, 2005).

Evaluation Versus Research

One way of measuring the value of a new program is to compare its results with those of a conventionally conducted course in the same subject. Most often this comparison cannot be made fairly because the two courses were planned to achieve entirely different objectives. It is very likely that there are no stated, measurable objectives for the conventional course that can be used as a basis for the comparison. Also, the subject matter treated in the two programs may be significantly different, with the content of the conventional course often being limited to a lower-level cognitive domain than that of the new program.

In some situations, evaluation is performed by using a formal research framework. This means that a carefully designed comparison study is based on control and experimental groups or classes. One or more hypotheses are stated as anticipated outcomes. Then, after instruction takes place, statistical methods are employed to gather data and report the evidence collected about learning outcomes. Conclusions are drawn that support or reject the initial hypotheses.

Such a methodology is usually more appropriate in basic or applied research studies that permit control over extraneous variables and allow for the establishment of reasonably equivalent experimental and control groups. Most instructional design projects are not planned to result in broadly applicable theories. Their purpose is to find out how well the needs that have been identified can be met. Growth in learner knowledge or skill activity, as measured by the difference between pretest and posttest results or by observing behavior before and after instruction, provides evidence of learning that can be directly attributable to the instructional program.

Sometimes the success in learning can be shown only in following up on-the-job work being done by individuals after instruction. For example, if after employees complete a safety course, accidents involving those employees are appreciably reduced (say, by more than 30%), then it can be inferred that the training was successful. Or, if company operating expenses decrease and revenues increase from

the pretraining to the posttraining period, then one could infer that direct benefits are a result of the training. On the other hand, when results do not meet goals, the evaluation evidence would indicate the shortcomings. Steps can then be taken to improve the program before its next use.

Thus, for evaluating instructional design projects, it is not necessary to perform formal research involving control/experimental groups and a detailed statistical analysis. All that must be done is to gather evidence relative to accomplishments or change from preinstruction to postinstruction for as many of the five components (i.e., effectiveness, efficiency, costs, attitudes, and benefits) as are considered important for that course, then interpret the information to reach conclusions about the success or failure of the instructional program.

A special note: For some of the procedures considered here, it is advisable (or even essential) to start collecting data at the time the program is initially planned. By doing this, you will have the necessary information to determine costs, time, and other facts pertinent to the evaluation.

PROGRAM EFFECTIVENESS

Effectiveness answers the question, "To what degree did students accomplish the learning objectives prescribed for each unit of the course?" Measurement of effectiveness can be ascertained from test scores, ratings of projects and performance, and records of observations of learners' behavior.

An analysis of scores can be prepared by hand or using a computer statistical package. The data may show the change from pretest to posttest results. Then, a summary may be presented in tabular form, as shown in Figure 13-3. The figure illustrates that the group, composed of six learners, accomplished 90% of the objectives. This figure is calculated by totaling the number of objectives satisfied (represented by the X marks in section c) and dividing by 6, which is the number of learners. The average number of objectives accomplished per learner is 4.5, which is 90% of the objectives. This result can be interpreted as a measure of the effectiveness of the instructional design plan for this group of learners. The percentage may be considered an effectiveness index representing the percentage of learners reaching a preset level of mastery (i.e., satisfying each objective) and the average percentage of objectives satisfied by all learners.

If all learners were to accomplish all objectives, the effectiveness of the program would be excellent. If 90% of the learners accomplish 90% of the objectives, can you report that the program has been effective? To answer this question, the instructor, along with the administrator or training director, must have previously decided the level at which the program would be accepted as effective. For a systematically planned academic course, attainment of the 80% level by at least 80% of the learners in a class could be acceptable as a highly effective program. In a vocational or skill area, 90–90 (90% of the trainees accomplishing 90% of the objectives) might be the accepted success level. Similar courses (e.g., in biology or electronics assembly) can be compared with respect to effectiveness indices and conclusions drawn for judging program effectiveness. Realistically, it is very likely

FIGURE 13-3
A sample analysis of test questions measuring cognitive objectives

a.

Unit Objectives	*Test Questions*
A	2, 4, 11
B	1, 7
C	3, 5, 12
D	8, 10
E	5, 9

b.

Learner	*Correct Answers to Questions*											
	1	2	3	4	5	6	7	8	9	10	11	12
AJ	x	x	x	x		x	x	x	x	x	x	
SF	x	x	x	x	x	x		x				
TY	x	x	x	x	x	x	x	x	x	x	x	
LM	x	x	x	x	x	x	x	x	x	x	x	
RW	x	x	x	x	x	x	x	x	x	x	x	
WB	x		x	x	x	x	x		x	x		x

c.

Learner	*Objectives Satisfied*				
	A	B	C	D	E
AJ	x	x	x	x	
SF		x	x	x	x
TY	x	x	x	x	x
LM	x	x	x	x	x
RW	x	x	x	x	x
WB		x	x	x	x

that, because of individual differences among learners and a designer's inability to design ideal learning experiences, no one can hope to reach the absolute standard of mastery or competency—100%—in all instructional situations. (Some training programs for which life and safety are critical—medical or airline pilot training, for instance—may require a 100% level of mastery.)

Then another question must be asked. Assume that your own performance standard requires all learners to accomplish 85% of the objectives but that as a group they actually satisfy 82% of them. Is the effort to reach the 85% level worth the cost? If not, you may have to settle for a somewhat lower level of accomplishment until someone can design a revision of the program that will enable reaching the desired level of performance with reasonable effort and cost. When evaluating the effectiveness of an instructional program, a designer must recognize that there may be intangible outcomes (often expressed as affective objectives) and long-term consequences that would become apparent only after the program is concluded and learners are at work. Both of these matters are given attention in the following sections as part of other summative evaluation components. Here, the evaluation of effectiveness is limited to those learning objectives that can be immediately measured.

Summative Evaluation Methods

The basic procedures for determining program effectiveness in summative evaluations are similar to those described earlier for formative evaluations (see Figure 13-1). Specifically, the major steps are these:

1. Specifying program objectives
2. Determining the evaluation design for each objective
   a. Pretest–posttest with one group
   b. One-group descriptive
   c. Experimental-control group
   d. Analysis of costs, resources, implementation
3. Developing data collection instruments and procedures for each objective
   a. Questionnaires
   b. Interviews
   c. Observations
   d. Achievement tests
4. Carrying out the evaluation
   a. Scheduling the data collection
   b. Collecting the data
5. Analyzing the results from each instrument
6. Interpreting the results
7. Disseminating the results and conclusions
   a. Evaluation report
   b. Group meetings
   c. Individual discussions

Data Collection Instruments. As with formative evaluations, data collection addresses one or more of the three domains of skills/behavior, cognitive, and affective. The main difference in summative evaluation is judging a completed rather than developing program.

For assessing skills, key information sources (as in formative evaluations) are as follows:

- Direct testing
- Analysis of naturally occurring events
- Direct/indirect observations
- Portfolios
- Exhibitions

For assessing cognition, measurement options include objective tests (e.g., multiple choice, true/false, matching) and constructed-response tests (e.g., short answer, essay, and problem solving).

Assessments of affective outcomes entail gathering reactions from both learners and the instructional staff as they look back on the program just completed.

FIGURE 13-4
Types of questions for gathering subjective responses

Checklist
Check each word that tells how you feel about the group projects and oral presentations used in this course.

_____Interesting	_____Informative	_____Difficult
_____Dull	_____Practical	_____Important
_____Exciting	_____Worthless	_____Stimulating
_____Boring	_____Useful	_____Unpleasant

Rating Scale
Compared with a typical lecture class, how useful was the format used in this experimental class for learning the course material? (Check one response.)
_____Better _____About the same _____Not as good
Now that you have completed the course, rate your feelings about history as a subject. (Circle the number that best reflects your reaction.)

Dislike very much	Dislike somewhat	Neutral	Like somewhat	Like very much
1	2	3	4	5

Rating
Please rank these topics as treated in the management course. Consider their value to you and your job. (Start with number 1 as the topic having the highest value.)

_____Planning	_____Organization and management
_____Self-assessment	_____Development
_____Stress	_____Personnel management
_____Labor relations	_____Performance appraisal
_____Effective presentations	_____Internal affairs management
_____Budgeting	_____Media relations
_____State-of-the-art technology	

Open-Ended Questions
What is your general reaction to this course: the objectives treated, the way it was conducted, your participation, its overall value to you, and so on?

Three categories of reactions may be given attention:

- **Opinions**: Judgments about the level of acceptance of course content, instructional methods, assistance from and relations with instructor and staff, study or work time required, grading procedure, and so forth
- **Interest**: Responses to the value of topics treated, learning activities preferred, and motivation for further study or work in the subject area
- **Attitude**: Reactions to the total program in terms of degree of its being pleasurable, worthwhile, and useful. Examples of types of questions for gathering subjective reactions are shown in Figure 13-4.

PROGRAM EFFICIENCY

In evaluating efficiency, three aspects of a program require attention:

- Time required for learners to achieve unit objectives

- Number of instructors and support staff members required for instruction and the time they devote to the program
- Use of facilities assigned to the program

Learner Time Required

Educational programs are designed typically in terms of available time periods—semesters, quarters, or other fixed time intervals (e.g., week, weekend, etc.). It is only when some flexibility is permitted that efficiency can be measured. If, for example, a conventional training program can be reduced from a period of possibly six to five weeks with the same or increased effectiveness in learning, the program can be considered efficient.

Efficiency can be used for measuring outcomes primarily of programs that give major emphasis to individualized or self-paced learning activities. From the learner's standpoint, the time required to satisfy unit or program objectives would be a measure of efficiency. Mathematically, this measurement is the ratio of the number of objectives a learner achieves compared with the time the learner takes to achieve them. Learners can be asked to keep records of time spent studying a unit or set of objectives (i.e., a time log). Or, in a more subjective fashion, an instructor can observe and make notations to indicate the number of learners at work in a study area during time periods.

For example, Mary satisfies seven objectives in 4.2 hours of study and work. Dividing the number of objectives Mary achieves by the amount of time it takes her to accomplish them yields her efficiency index, 1.7 (7/4.2). Bill achieves the seven objectives in 5.4 hours. His efficiency index is therefore 1.3. Thus, the higher the index, the more efficient the learning. Such an index can be calculated for each unit then averaged for each learner to give an efficiency index for the course. Keep in mind that many instructional programs will not yield such easily attainable and concrete measures of mastery over time. Also, efficiency indices may not be comparable across different units of instruction because of the nature of the material taught and the characteristics of the students. However, where feasible, an efficiency index or some other quantitative measure can provide highly useful information for evaluating allocations of time and resources. A clear advantage, especially in business contexts, is that quantitative measures often are expected and given more credibility by stakeholders than are subjective types of evidence.

Faculty and Staff Required

The number of faculty and staff positions required for instruction, supervision, or support of an instructional program also relates to efficiency. The question is, "How many learners are being served by the staff?" If a course requires a half-time faculty position plus the equivalent of one full-time position in assistants and technicians to serve 48 learners, then the faculty-to-learner ratio would be 1:32 (i.e., 1.5:48. If the institution-wide ratio of faculty-to-student load is 1:20, then the lower ratio of 1:32 indicates a more efficient use of faculty and staff personnel. Greater efficiency, however, may not necessarily mean greater effectiveness.

The ratio of 1:32 may be reported on paper, but the actual working time of faculty and staff in the program can give another indication of efficiency. Let's assume that the same instructor and support staff (i.e., the 1.5 positions) are spending 60 hours a week on the program doing preparation, teaching, consulting with learners, evaluating performance, marking tests, providing resources, etc. If normal time devoted to a course is 45 hours per week for a staff of similar size, then the procedures may need some revision.

Use of Facilities

Another factor of efficiency is the amount of time that learning facilities—classrooms, learning labs, and so forth—are available during a day, a week, or other time period. If a facility is used 12 hours a day, this may be considered an efficient use of space. By obtaining these data as a program is expanded, the need to increase use or to provide for additional training space can be evaluated.

A second component of efficient space utilization is the number of learners using the facility during a time period. When 110 learners are being served in a 15-station microcomputer lab on a weekly basis, this may be seen as an efficient use of space. Keep records so that the time learners and staff spend in the program and in the facility can be calculated and objectively related to this factor of efficiency.

PROGRAM COSTS

Historically, a major concern in educational programs is the cost of instruction. Expense categories, such as personnel, equipment, and supplies, are established to aid the administration in controlling and reporting about programs. Standard bases that are frequently used for allocating funds in educational budgets are average daily attendance (ADA) in public schools, full-time equivalent (FTE) in higher education (i.e., the total number of courses being taken divided by the number of courses in a full course load), the number of faculty assigned in terms of FTE, and student credit hours or student-to-faculty contact hours. These bases for allocating funds are mainly accounting methods. They provide little information about the real costs of a single program.

Although a school or college is not the same as a business operation, for both, specific factors affecting costs can be identifed and controlled. The education and training literature contains numerous explanations and reports on how program costs can be derived. Formulas that consider many of the variables that affect costs are presented in detail, and their complexities are interpreted. Such terms as *cost*

effectiveness, cost efficiency, and *cost benefits* are frequently used. Our concern here is simply to answer the question, "What does it cost to develop and operate a specific program for the number of learners served?" Once we have this essential information, we are able to relate costs to effectiveness, efficiency, and resulting benefits; thus, we are able to judge the acceptability of program costs. Any new course or a program being revised requires attention to the two major categories of costs: developmental and operational costs.

Developmental Costs

As an instructional project is being planned and developed, some or all of the following costs, sometimes called start-up costs, may be incurred:

- Planning time: Amount of salary paid out for time spent by each member of the planning team on the project, calculated from the percentage of total work hours spent on it or the number of hours spent by each member multiplied by his or her hourly or monthly salary rate, plus fees for consultants
- Staff time: Amount of salary paid for time spent by each member engaged in planning, producing, and gathering materials, calculated from the percentage of total work hours spent on it or the number of hours spent by each person, multiplied by his or her hourly salary rate
- Supplies and materials for preparing print, media, and other materials
- Outside services for producing or purchasing materials
- Construction or renovation of facilities
- Equipment purchased for instructional uses
- Expenses for installing equipment
- Testing, redesign, and final reproduction of resources in sufficient quantity for operational uses (includes personnel time and costs of materials and services)
- Orientation and training of personnel who will conduct instruction
- Indirect costs: Personnel benefits such as retirement and insurance, related to time and salary charged to the project (this information is typically available from the personnel department)
- Overhead: Utilities, furniture, room and building costs or depreciation allowance, proportion of other institutional services charged to the project (this information is usually available from the business manager or the controller of the organization)
- Miscellaneous (e.g., office supplies, telephone, travel, etc.)

Here is an example of the developmental costs for a general education college-level course involving two instructors. It includes large-group presentations incorporating PowerPoint presentations, student self-directed learning with 10 interactive multimedia units, a study guide, and small discussion sessions.

Design Time	
2 instructors, 1 month summer	$16,500
Instructional designer, 1 month summer	$9,500
Graduate assistant, 100 hours	$2,500
	$28,500
Development Time	
2 instructors, 0.25 time, 1 semester	$22,000
Instructional designer, 0.25 time, 1 semester	$14,250
Programming	$22,000
Graphic artist, 80 hours	$3,500
	$59,750
Materials and Supplies	
Office supplies	$250
Equipment	
20 computers for lab	$24,000
Renovative Facility	
20 learning stations	$16,000
Electrical wiring	$9,000
	$25,000
Other Costs	
Formative evaluation and revision costs	$12,000
Staff benefits	$14,600
	$26,600
Total Development Costs	**$164,100**

Operational Costs

When the project is fully implemented and instruction is taking place, the recurring operational costs include the following:

- Administrative salaries (based on percentage of time devoted to project)
- Faculty salaries for time spent in the program (e.g., contact hours with groups and individual learners, planning activities, evaluating program, revising activities and materials, personnel benefits)
- Learner or trainee costs (applicable in business-oriented training programs: e.g., salary, travel and lodging, income for company reduced while trainee is not on job, or replacement cost of a person substituting for a trainee job)
- Salaries for assistants, maintenance technicians, and others
- Rental charges for classroom or other facilities if offered at an off-campus location

- Replacement of consumable and damaged materials
- Repair and maintenance of equipment
- Depreciation of equipment
- Overhead (e.g., utilities, facilities, furnishings, custodial services, etc.)
- Evaluation and update of materials (i.e., time and materials)

Here is an example of the operational costs for the college-level course shown in the previous example over a one-semester term:

Salaries	
2 instructors, 0.25 time	$22,000
Benefits	$5,000
2 graduate assistants	$22,000
Subtotal	$48,000
Replacements and repairs	$2,500
Total Operating Costs	**$50,500**

Instructional Cost Index

We cannot attempt to judge whether the costs of an instructional program are acceptable by looking solely at the gross amount expended. If it costs a company $1,000 to manufacture pencils, this sum must be related to the number of pencils made. Then the price per pencil has meaning and can be compared with the price per unit manufactured by other companies. In an instructional program, costs should be related to the number of learners served in the program.

With data available on developmental and operational costs, we can calculate the cost per learner for a program. This is the important bottom-line amount that allows for comparison of costs between programs, leading to the acceptance of expense levels. Cost per learner or trainee may be labeled an instructional cost index. It is determined by the following procedure:

1. Spread the developmental costs over a series of time periods (e.g., 10 training sessions or 5 semesters). This would be the anticipated life of the program before it should require major revisions or cease to be useful. This procedure is known as amortizing the cost.
2. Add together the preceding prorated amount of the developmental costs (for 3 years) and the operational costs for one use period (either a complete training class or an academic semester).
3. Determine the average number of learners known or anticipated to be in the program with each use. Divide the total in step 2 by this number. The result is the cost per learner, or the instructional cost index.

An example of an instructional cost index calculated from the previous example of developmental and operational costs follows:

Total operational costs	$50,500
Portion of developmental cost (164,100/6)	$27,350
Total cost per semester	$77,850
Number of learners in program	340
Instructional cost index (77,850/340)	$228.97

(This is the total cost for each learner over one semester.)

If this program continues beyond five semesters (at which time all developmental costs will have been amortized) and the number of learners remains the same, the instructional cost index will then drop to 148.53 (50,500/340). During this period, limited funds are included for minor updates and revision of materials. At the end of five semesters, a reexamination of the program for this course may be advisable. The course then may be continued as is, or new developmental costs—hopefully lower than the original ones—would be required. These would affect the ongoing instructional cost index.

The index number itself has little meaning. Calculations could be made in the same way for traditional program costs in a comparable training or subject area. As previously stated, it is difficult (and usually unfair) to make a comparison between a new program with carefully structured objectives and a traditional program based on generalized objectives. It would seem more appropriate to compare two skill-type training programs, two math classes, or a biology and chemistry course if each one has been systematically planned and implemented. Once an instructional cost index has been calculated, the instructor or designer should ask these questions:

- Is the program cost-effective? This is a subjective decision, but useful information can be obtained by relating the instructional cost index to the level of learning outcomes (e.g., 90% of the learners accomplish 84% of the objectives). If a satisfactory learning level is reached and the instructional cost index seems to be within reason, the program would be considered cost-effective.
- Is the program cost-efficient? Relate the instructional cost index to efficiency factors (time required by learners to complete activities, staff time required for instruction and support, level of facilities' use). If the efficiency index seems acceptable, with a reasonable instructional cost index the program would be cost-efficient.
- Are the costs justified in terms of resulting benefits (cost–benefit analysis)? Relate the instructional cost index to the benefits that a company or other organization derives from personnel who complete the training program. (See the following section for details and discussion of potential benefits resulting from training.) If the benefits are high and costs acceptable, then the answer to this question is yes.

If the outcomes of a program prove to be acceptable but the instructional cost index remains higher than desired, certain steps might be taken to lower the operational cost portion of the index, as follows:

1. Consider the feasibility of including more learners in the program (as in a distance-learning course using television). Perhaps more individuals can be served without reducing the quality of the instruction.
2. Decide whether assistants might replace instructors for certain activities without lowering the effectiveness of the program. This would reduce higher-cost instructor time.
3. Plan to relieve instructors of some learner contact time by developing additional self-paced learning activities for learners.
4. As a last resort, reduce the training time or lower some of the required performance standards. Shorter instructional time would reduce instructor time and thus costs.

An alternative cost index measure is to calculate the index in terms of total contact hours. Thus, a weeklong course for 20 people has 800 total contact hours (40 hours × 20 people).

CONFIRMATIVE EVALUATION: DETERMINING OUTCOMES OVER TIME

Instructional programs are most often offered for three general reasons:

- To "educate" individuals so that they may participate as informed, cultured, and productive citizens in society
- To prepare individuals for a gainful vocation
- To improve or upgrade competencies of individuals in a specific task or in certain aspects of a job

For each of these reasons, determining the success of an instructional program requires attention to important outcomes beyond the results of written and performance tests given at the end of a unit or a course. Often the accomplishment of major goals or terminal objectives stated for a program can be assessed only some time after instruction is concluded. As emphasized throughout this chapter and the preceding ones, the evaluation of instruction needs to be continuous. In our view, distinguishing between the constructs of formative, summative, and confirmative evaluations is mainly important in considering what the evaluation is likely to emphasize and when in the instructional design and implementation process the evaluation will be conducted. From an operational standpoint (e.g., planning, instrumentation, data collection, and analysis) the three evaluation types are much more similar than they are different. Confirmative evaluation represents a continuation of summative evaluation. Both approaches are designed to judge the effectiveness of instruction: summative soon after instruction is completed and confirmative after some time has passed.

Our practical view of both summative and confirmative evaluation is that rarely will either be used solely to pronounce an instructional program as "working" or "not working" and then conclude. In the real world, there will typically be opportunity to use the results of these evaluations to make improvements in future training. Thus, the important aspects of formative evaluation continue to be employed, and the instructional program regarded as never fully "completed."

Confirmative evaluation may encounter a few hurdles that earlier evaluations do not. After a course is completed, the learners or trainees move to other courses or work at different locations. Observations of them at work or communication with them may require an extra effort. Some important outcomes are in the affective domain. These may be difficult to identify and measure. Responses for evaluation may be needed from other persons (colleagues, supervisors, and others) who may not be understanding or cooperative. Regardless of these obstacles, attempts should be made to follow up on learners after an instructional program has ended. Evidence of follow-up benefits could be the most important summative results to measure.

Approaches to Confirmative Evaluation

There are two basic types of situation that warrant conducting confirmative evaluation studies. One type is *learner oriented* and concerns the degree to which, as time passes, consumers of the instruction retain the skills and knowledge needed to perform at desired levels. A second type is *context oriented* and concerns the degree to which the instructional product remains effective as conditions (e.g., policies, politics, resources, technological advances) change over time.

Learner-Oriented Approach. Several years ago, two of us were hired by a large chemical corporation to evaluate its employee training courses in areas such as communication, public speaking, and interpersonal relations. Although the corporation routinely conducted follow-up evaluations of employee skills and attitudes soon after participants completed the courses, there was strong interest in assessing longer-term impacts. The corporation wondered whether, for example, although the employees might do well on an immediate performance or knowledge test after completing the course on making oral presentations, would they demonstrate the desired skills on the job a year later? If not, the need to strengthen the original course or offer refresher training would be implied. The confirmative evaluation that we conducted involved asking managers, peers, and subordinates to rate former trainees' on-the-job performance on target skills. A second measure was asking the trainees to rate retrospectively the helpfulness of various items of course content and activities for fulfilling their job requirements. From these data, the evaluation study identified course components that were successful for long-term achievement of objectives and those that were not.

Note that the measures employed in this example and in many confirmative evaluations were fairly low-cost and coarse-grained indicators of continued program success. Had the results indicated major weaknesses in these continuing training courses, refinements in the course designs accompanied by a more granular formative evaluation would have been implied.

Context-Oriented Approach. But what if the desired objectives of instruction or training change over time? Obviously, a course or unit that was formerly judged

effective (in formative and summative evaluations) would no longer be optimum. The fact that learners might retain the skills taught over long periods of time would hardly matter if the skills mastered were no longer the ones desired. A second type of confirmative evaluation examines whether the instruction is achieving objectives following changes in conditions and policies as time passes.

For the most part, the context-oriented confirmative evaluation would use the methodology of summative evaluations. Although stakeholders in the instruction would hope for continued supportive results, the evaluation may uncover that the program's effectiveness has diminished due to changes in

- Learner characteristics
- Curriculum or performance expectations
- Technology
- Budgetary support for the program
- The training need
- Teacher support or preparation for using the program

Educational Programs

Traditionally, the general, long-term benefits of educational programs are measured through statewide and national standardized tests given to students in public schools, undergraduate and graduate admission examinations for college students, and regional or national opinion surveys conducted at various times. Such tests measure broad, fairly general objectives. One limitation is that gains on these tests may not show up immediately after a new instructional program is implemented. The summative evaluation, therefore, may not indicate success, but a year later, a confirmative evaluation might. A second limitation, which affects both evaluation approaches, is that the objectives measured may be too broad to provide useful information on how well a particular course accomplished its specific instructional objectives. Concerns may also arise about the circumstances under which students prepare for or complete the standardized tests (Haladyna, Nolan, & Haas, 1991).

The long-term outcomes of the objectives of specific courses are frequently not examined or examined only casually. Within the framework of the goals and terminal objectives of a program, the following categories of outcomes might be considered:

- Capabilities in basic skills (e.g., reading, writing, verbal expression, and mathematics) required in following courses
- Knowledge and competencies in a subject as bases for study in subsequent courses
- Proficiencies to carry out job tasks and responsibilities in occupational employment
- Fulfillment of role as good citizen (e.g., law abiding, participating in democratic process, etc.)

As indicated at the beginning of this section, data concerning these outcomes are not easy to obtain. The following methods are commonly used in continuing (i.e., confirmative) evaluations to gather information:

- Completing questionnaires: Ask former students, present instructors, or employers to respond to a questionnaire designed to indicate learners' present proficiencies related to competencies derived from the course or program that is being evaluated (see Figure 12-2).
- Conducting interviews: Meet with former learners, present instructors, or employers to inquire about the present proficiencies of learners as related to competencies from the course or program being evaluated.
- Making observations: Observe learners in new learning or performance situations and judge their capabilities as a follow-up on competencies acquired in the course being evaluated.
- Examining records: Check grades and anecdotal records of students in school files to ascertain how they are now performing in their classes related to competencies gained in the course being evaluated. (Note: Because of privacy laws, this procedure may require permission from the former students before records can be retrieved.)

Training Programs

A training program within a business concern, an industrial company, a health agency, or other organization usually has clearly defined outcomes to be accomplished. These planned results may have been identified initially when a needs assessment (see Chapter 2) was first made. The consequent benefits are expected to result in improved job performance and often can be translated into dollar savings or increased income for the company.

Three areas may need to be assessed in posttraining evaluation: appropriateness of the training, competencies of the employees, and benefits to the organization.

Appropriateness of the Training. Although the program was developed according to identified needs, changes in on-the-job operating procedures and the equipment used could necessitate job performance different from what was taught. Use confirmative evaluation to determine whether modifications are required before training is conducted the next time.

Competencies of Employees. It is one thing to pass written tests and perform satisfactorily in the controlled environment of a classroom or laboratory but, potentially, another to be successful in transferring the learning to a job situation. Determine how well the former trainees now perform the job or tasks they were trained to do.

Benefits to the Organization. The advantages need to be measured in terms of the payoff to the organization as well as to the individual. Some of the criteria that indicate that a training program has been beneficial to the organization are these:

- Reduced number of accidents through increased safety
- Increased service abilities, including both work quality and performance speed

- Improved quality of products being produced
- Increased rate of work or production
- Reduced problems with equipment due to malfunctions and breakdowns
- Increased sales of products and greater services, or more income being generated (referred to as "return on training investment")

With respect to affective-type outcomes, the following may be some of the expected results:

- Less employee tardiness and absenteeism
- Less employee turnover
- Greater job satisfaction
- Higher level of motivation and willingness to assume responsibilities
- Increased respect for the organization

The same methods for gathering information for educational programs would apply to measuring the follow-up benefits of a training program: questionnaires, interviews, observations, and examining records. In terms of actual performance levels, if careful records are kept, comparisons can be made between pretraining and posttraining competencies. A key method of follow-up evaluation can be related to measuring effects on company expenses or revenue. This approach requires a comparison of pretraining cost factors with the costs and income data determined at a reasonable time after training is completed. This evidence can be one of the best measures to relate training benefits to the bottom line, with which a company is most concerned. Keep in mind, though, that not all training courses will have a direct, measurable impact on the bottom line (e.g., courses in interpersonal relations, public speaking).

In summary, confirmative evaluations take up where summative evaluations leave off—once the instruction is completed, the application of learning begins and continues over time. A confirmative evaluation will tend to have the following characteristics:

- Continuous (i.e., repeated over time, where feasible)
- Occurs in realistic contexts (i.e., on the job, in practice)
- Emphasizes performance rather than simply knowledge
- Includes learner reactions as direction for improvement
- May address new evaluation questions as performance requirements or contexts change over time
- May use smaller samples (more case studies) due to attrition of original learner cohort
- May use indirect measures (e.g., colleague, supervisor, and self-ratings; data from naturally occurring events) because of the difficulty of collecting follow-up data in the field

REPORTING RESULTS OF SUMMATIVE AND CONFIRMATIVE EVALUATIONS

The final step in creating evaluations of instruction is to prepare a report of the results for others to read and examine. Careful attention should be given to this activity. Future support for the program, as well as the assistance required for additional instructional design projects, can be influenced by the manner in which a summative or confirmative evaluation is reported. First, the evaluator must decide for whom the report is to be prepared—administrators/training managers, instructors, or another supporting agency. By considering those persons who are to receive the report, emphasis or special attention may have to be given to certain phases of the evaluation. For example, how and where funds have been spent may be of primary interest, or evidence of follow-up benefits may be of more value to the readers than are the efficiencies or effectiveness of instruction.

Second, the evaluator must decide on the format of the report. Should it be on paper for individual reading, or will it be presented to a group with the support of slides or overhead transparencies? In either case, plan to report results attractively. Not everyone will be as highly interested or as well informed about the project as the designer has been. Here are some suggestions:

- Give the report an interesting title.
- Summarize highlights so the key outcomes can be grasped quickly. Do this by setting them off on a page with white space or boxing each statement.
- Describe supporting data in visual ways with graphs rather than as detailed tables; use artwork as appropriate.
- If slides or PowerPoint will be prepared, limit the information displayed to only the key points. Prepare printed materials that correlate with the visuals and contain the details of information for the audience to retain.
- End by making appropriate recommendations for continuing, extending, modifying, or terminating the program.
- Where feasible, adapt the style and content of the report to the main target audiences. Different stakeholder groups will have different backgrounds, interests, and expectancies for what the report will convey.
- The reporting format for formative evaluations outlined in the section "Reporting" is generally appropriate as well for summative and confirmative evaluations.

SUMMARY

1. Formative, summative, and confirmative evaluations serve the complementary purposes of assessing both developing and completed instructional programs.
2. A basic model for planning formative evaluations addresses eight areas: purposes, audience, issues, resources, evidence, data-gathering techniques, analyses, and reporting.

3. Common formative evaluation approaches can be classified as (1) connoisseur based, in which expert opinions are sought; (2) decision oriented, in which information related to particular questions is gathered; (3) objectives based, in which assessments are made of the degree to which particular objectives are obtained; and (4) public relations inspired, in which financial support or backing for a project is solicited based on the evaluation findings.
4. Recently proposed *constructivist-oriented* formative evaluation approaches focus on three factors related to learners' interactions with computer-based instructional support: *learning constructs* (interactivity, feedback, graphics), *quality of instruction* (help supports, instructions, organization), and *learner engagement* (interest, motivation, enjoyment).
5. Three major stages of formative evaluation consist of one-to-one trials, small-group trials, and field trials. Each successive stage focuses on a more developed version of the instructional program, using larger samples of students.
6. Summative evaluations, unlike experimental research, are designed to provide information about specific instructional programs. They are not used to test general theories and thus do not require rigorous research methods or control groups.
7. Program effectiveness is determined by analyzing test scores, rating projects and performance, and observing learner behavior.
8. Summative evaluations involve similar planning procedures and methodologies as do formative evaluations. The emphasis in summative studies, however, is on the full, completed program rather than on preliminary versions of the program.
9. A useful outcome measure for summative evaluations is program efficiency, computed as a ratio of the number of objectives achieved to the time taken to achieve them.
10. Program costs are evaluated by determining both developmental costs (the expense of designing the program) and operational costs (the expense of offering the program). The instructional cost index is the total cost per learner or trainee.
11. Confirmative (follow-up) evaluations should be conducted after the student leaves the program. In the case of educational programs, the important outcomes are basic skills, knowledge, competencies, and performance in occupational settings. For training programs, important outcomes include appropriateness of the training, competencies of employees, and benefits to the organization.
12. *Learner-oriented*, confirmative evaluations concern the degree to which, as time passes, consumers of the instruction retain the skills and knowledge needed to perform at desired levels. *Context-oriented* studies concern the degree to which the instructional product remains effective as conditions change over time.
13. The evaluation report should be carefully prepared and attractive in its appearance/presentation. Adaptation of content to stakeholder backgrounds and interests is highly important.

THE ID PROCESS

Inexperienced designers sometimes view formative and summative evaluations as discrete strategies that respectively come at the "beginning" and at the "completion" of the design process. Our approach is to see them, along with confirmative evaluations, as overlapping processes having more similarities than differences. Although every design project is different in content, context, and goals, we offer the following as general guidelines based on our experiences.

Make continuous evaluation a primary component of the initial design plan. Once a project is launched, it seems to be human nature to want to devote more time and money to designing and implementing the instructional program. It is generally difficult to add "more" evaluation if it is not built in at the front end. In most contexts, being sensitive to the political aspects of a design project is critical to the project's success. Of the many stakeholders concerned (e.g., sponsors, learners, administrators, and instructors), there are likely to be both supporters and detractors. Evaluation, therefore, not only becomes a tool for refinement and judgment, it becomes a valuable means of communicating (with data) what is being accomplished. We have discovered many times that "data talk," particularly when there is disharmony among stakeholders on such matters as the need for the course, its content, how and when it is offered, the costs involved, or the delivery methods used.

Although the distinctions between the three types of evaluation is useful as a heuristic for evaluation timing, emphasis, and orientation, it seems most important to employ each as the evaluation objectives and conditions dictate at each point in time. There will be times when a formative evaluation that is conducted very early in the design process must produce preliminary data to satisfy the new training manager who wants some indication of the project's impact to support it. Such situations are not a matter of how the "book" says a pure formative evaluation must be done. Simply put, unless that particular manager receives some outcome data, there will be no project on which to conduct a pure formative evaluation! And there will be situations (probably many) when it makes perfect sense in a summative or confirmative evaluation to gather information for improving the design. Judging from our experiences, the art and science of instructional evaluation in the real world often necessitates developing well-constructed hybrid designs adapted to situational needs.

A final consideration in all phases of evaluation is to what degree outside experts should be employed. Different types of experts can be particularly valuable. The SME (e.g., a database administrator consulting on a course on database management) should be helpful in validating the content of the instruction. A media expert may be needed to help with instructional delivery using specialized technology (e.g., distance learning) with which the designer is less familiar. An external, professional evaluator may be needed to design the study, develop instruments, and analyze and interpret the data. When funding permits, we strongly recommend using experts. Not only is their knowledge important, but what they say adds objectivity and credibility to the project. Expert data usually "talk" very convincingly.

APPLICATIONS

You are designing an instructional program to be used at kiosks in an art museum to acquaint visitors with the museum layout and exhibits. It is important to the philanthropic foundation that supported the stations that visitors (1) actually use the program and (2) find the information useful and interesting. Using examples, describe how each of the following types of formative evaluation might be used: connoisseur based, decision oriented, objectives based, and public relations inspired. Which of these orientations might you also apply in summative and confirmative evaluations?

ANSWERS

Assuming that you are not highly knowledgeable about the particular art museum or art exhibits in general, connoisseur-based evaluation would be useful to confirm the accuracy of the content of the program, its organization, and its presentation. Decision-making evaluation might focus on questions such as, "Is the program economical to update as new exhibits are featured?" Answers to these questions will provide useful suggestions for refining the design. Objectives-based formative evaluation might focus on the criterion questions posed by the sponsoring foundation (as well as other relevant outcomes): "What percentage of visitors use the program and for how long?" and "Do users find the program to be helpful and interesting?" Given that this evaluation is formative, the results will be used less to judge its effectiveness than to provide directions for its improvement. For example, infrequent use of the program may lead to the decision to place the information stations in more visible locations in the museum lobby. Public-relations-inspired evaluation may involve providing the foundation with preliminary information showing that good progress is being made in the program design and implementation.

For summative and confirmative evaluations, the two most important orientations will probably be objectives based, to determine how effectively the completed program is achieving the desired outcomes, and public relations inspired, to keep the foundation satisfied and perhaps (assuming the program is successful) attract additional funding for related projects. With a completed program, connoisseur-based evaluation would probably not be needed. Assuming that the "completed" program can still be improved, decision-based evaluation, however, might be a helpful supplement providing data on how effectively different program components are working.

INSTRUCTIONAL DESIGN: DECISIONS AND CHOICES

The draft materials for the unit on using a fire extinguisher have been completed and are ready for formative evaluation. You devise the following evaluation strategy:

1. SME review for content accuracy
2. Individual learner review of materials

3. Classroom instructor feedback
4. Classroom and Web test-item analysis

After the SME has signed off on accuracy of the content, you ask the client to provide three learners (ideally individuals from the primary audience such as mailroom clerks) to try out the instruction. You greet the first test subject and sit him down in front of a computer workstation. You tell the mailroom clerk, "We want to make sure that these instructional materials are effective, so your feedback is important to us. As you go through these materials on the Web (screen pages of instruction and practice activities), tell me what you are thinking, particularly if something you read is hard to understand or confusing, or if something works particularly well."

The learner then proceeds to go through the Web-based materials. [Some designers prefer to test the instruction before the actual Web pages are developed. In this case, a draft is prepared on paper, in the form of sketches of screen shots or as PowerPoint slides with links, and these are provided to the learner.] While the learner goes through the instruction, you observe, particularly noting where the learner pauses, returns to a previous screen, or rereads a section. You also look for a confused facial expression or other signs that the learner is stuck. After the learner completes the instruction, you interview him. Here are some sample questions:

1. What did you think of the instruction?
2. Was anything difficult to understand?
3. Could you relate to the examples?
4. How did you feel about the practice activities and feedback?
5. Are there any changes you would recommend?

[Notice how the first question is open-ended and is intended to obtain data about the learner's overall reaction to the instruction. The second question is specifically aimed at discovering weaknesses in the instruction. The third question will tell the designer whether the examples are good or need to be replaced. (The designer will want feedback from several test subjects before deciding to change any of the examples.) The fourth question, along with the posttest results, tells the designer whether the practice activities and feedback are effective.]

You revise the materials based on the feedback from the first test subject. These "fixes" may include revising sentences, changing the sequence of information, changing the practice activities, or correcting spelling. Sometimes, you may not be certain that a change suggested by the test subject is warranted. In this case, you hold off making that change until you receive feedback from a second or even third test subject.

You have also decided to sit in on the first classroom offering. You observe the learners taking notes, and you carefully attend to the questions that the learners ask the instructor. You also watch the learners practice using the fire extinguisher and note where the instructor provides new information or must review procedures. After the instruction you interview all the participants as a group for about 10 minutes (essentially a focus group). In addition to the questions used for the one-on-one tryouts, you add the following:

- How did you feel about the hands-on practice?
- How did you feel about the pacing of the instruction?
- Did you have enough time?

Next, you debrief the classroom instructor using a set of questions that you have prepared in advance. You ask questions such as:

- What is your overall impression of the instruction?
- Did the learners have any difficulty with the material?
- Were the in-class practice activities effective?
- Are there any changes that should be made to the performance checklist?
- Are there any other changes that we should consider?

Last, you examine the posttest results and prepare a summary chart:

Question #	*Objective #*	*% Correct/Accuracy*
1	1	40%
2	2	85%
3	2	80%
4	3	100%
5	4	95%
6	4	100%
7	5	95%
8	6	100%
9	6	100%

Based on the posttest data, the introductory conceptual material appears to need revision. But, you also hypothesize that question 1, rather than the instruction, might need to be revised. After examining all the data, you conclude that both the instruction and the test item need revision.

REFERENCES

Brenneman, J. (1989). When you can't use a crowd: Single subject testing. *Performance and Instruction, 28*, 22–25.

Clarebout, G., & Elen, J. M. (2008, March). *Help-seeking and advice: A complex relationship?* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Dick, W., & Carey, L. (1991). *The systematic design of instruction* (2nd ed.). New York: Harper Collins.

Flagg, B. N. (Ed.). (1990). *Formative evaluation for educational technologies*. Hillsdale, NJ: Erlbaum.

Friesen, N., & Anderson, T. (2004). Interaction for lifelong learning. *British Journal of Educational Technology, 35*(6), 679–687.

Gooler, D. D. (1980). Formative evaluation strategies for major instructional development projects. *Journal of Instructional Development, 3*, 7–11.

Haladyna, T. M., Nolan, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the analysis of test score pollution. *Educational Researcher, 20*, 2–7.

Kay, R. H., & Knaack, L. (2009). Assessing learning quality and engagement in learning objects: The Learning Object Evaluation Scale for Students (LOES-S). *Educational Technology Research and Development*, 57(2), 147–168.

Krauss, F., & Ally, M. (2005). A study of the design and evaluation of a learning object and implications for content development. *Interdisciplinary Journal of Knowledge and Learning Objects, 1*. Retrieved from http://ijklo.org/Volume1/v1p001-022Krauss.pdf

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Nurmi, S., & Jaakkola, T. (2006). Effectiveness of learning objects in various instructional settings. *Learning, Media, and Technology*, 31(3), 233–247.

Ross, S. M. (2005). Assessment: Summative evaluation. In V. A. Anfara, Jr., G. Andrews, & S. Mertens (Eds.), *The encyclopedia of middle grades education* (pp. 126–128). Greenwich, CT: Information Age Publishing.

Ross, S. M., & Morrison, G. R. (1995). Evaluation as a tool for research and development. In R. D. Tennyson & A. Barron (Eds.), *Automating instructional design: Computer-based development and delivery tools* (pp. 491–522). Berlin: Springer-Verlag.

Saab, N., Gijlers, H., Van Joolingen, W. R., & van Hout-Wolters, W. R. (2008, March). *Using or not using the provided tool during collaborative inquiry learning*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Sterbinsky, A., & Ross, S. M. (2005). Assessment: Formative evaluation. In V. A. Anfara, Jr., G. Andrews, & S. Mertens (Eds.), *The encyclopedia of middle grades education* (pp. 122–126). Greenwich, CT: Information Age Publishing.

Thiagarjan, S., Semmel, D., & Semmel, M. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*. Bloomington, IN: Center for Innovation in Teaching the Handicapped.