

Domain of Evaluation

In the world of instructional technology, evaluation can mean measuring the worth of a particular training program or module. It can also mean determining the appropriateness or effectiveness of training materials. Lastly, evaluation can mean identifying the changes that can improve learning and instructional materials. It is sometimes a process of using potential learners to test instructional materials and strategies to inform the revision process.

Seels & Richey (1994) define evaluation as “determining the adequacy of instruction and learning.” (p.53) Evaluation processes are found in all stages of the process of instructional technology. Furthermore, comprehensive models exist that help guide the instructional technologist in planning for evaluation. These models are explained in more detail in the following section. Instructional technologists use evaluation techniques to perform needs assessments (problem analysis), develop assessments based on defined objectives (criterion-referenced assessment), determine revisions that need to be made during materials development (formative evaluation) and identify the strengths and weaknesses of an instructional program following its full implementation (summative evaluation). These four components of evaluation form the subcomponents of the definition of the domain of evaluation and are distinguished by the types of activities they measure (Seels & Richey, 1994).

Problem Analysis is the process of utilizing information-gathering and decision-making strategies to identify a problem. The ultimate goal of problem analysis is to determine goals and priorities by identifying needs (Seels & Richey, 1994). A need is best defined as a gap between what is currently happening and what should be occurring in terms of results (Kaufman, 1972, Rossett, 1987).

Assessing the needs of an organization based on an identified problem is the first step to identifying causes and developing solutions, or interventions. The process of needs assessment is a systematic process of using surveys, interviews and observations to determine what is and what should be by gathering information about the actuals, optimal and the feelings surrounding the problem (Rossett, 1987). The products of this type of evaluation are the identification of a problem and its causes, an understanding of the nature of the problem described in actual and optimal performance, and an identification of “gaps” (Rossett, 1987). Gap identification is the first step to developing instructional goals and objectives.

Criterion Referenced Assessment is one of two basic ways of Interpreting student performance. The other is norm-referenced assessment, a technique used to measure performance in relation to other learners. Criterion-referenced assessment is the process of utilizing techniques to determine learner mastery of specific content objectives (Seels & Richey, 1994). A criterion-referenced assessment is designed as a measure of performance that can be interpreted in terms of learning tasks that are defined and categorized by domains (Linn & Miller, 2005, p. 37).

As an evaluation of learning, criterion-referenced assessment offers the technologist a way to judge whether or not learning actually occurred through activities carefully aligned to a certain set of objectives. If designed effectively, a criterion-referenced assessment can describe the specific knowledge and skills each learner can demonstrate. The instructional goals and objectives generated as a result of the analysis serve as the foundation for assessment. These goals serve as guidelines for assessment by defining the intended learning outcome so that appropriate activities are developed. The process of developing assessment activities involves ensuring the alignment, validity, reliability and

usability in order to evaluate learner performance using methods that measure the objectives in an appropriate, consistent and practical manner (Linn & Miller, 2005).

Formative Evaluation is most often used in the early stages of instructional development to gather information of the adequacy of materials, strategies and assessments. This information is used in the revision and further development process (Seels & Richey, 1994). The use of formative evaluation techniques can be used to improve the effectiveness of instruction and should not only be used after the first draft of instruction is developed, but as a process of testing assumptions about instructional techniques throughout the processes of design and development (Dick, Carey & Carey, 2001). The focus of this type of evaluation is not limited to instructional materials, but is also used in other areas of the ISD process.

Formative evaluation can also focus on examining and changing processes as they are happening (Boulmetis & Dutwin, 2005). For example, an instructional technologist may evaluate how well a specific process in program implementation is working by conducting a formative evaluation. Rather than an evaluation of the development process, the technologist is judging the effectiveness and efficiency of the implementation process (Boulmetis & Dutwin, 2005). Formative evaluation can be used to analyze instructional products or programs in terms of participant reaction, evidence of learning, evidence of organizational support, participant use of new knowledge, or in collecting initial evidence of intended learning outcomes (Guskey, 2000). Use of this data by an instructional technologist facilitates program, material or product revision.

Summative Evaluation focuses on the outcomes of an instructional program or performance improvement intervention. It answers the question, “Has the problem been solved?” (Seels & Glasgow, 2001, p. 313). It “involves gathering information on adequacy and using the information to make utilization decisions” (Seels & Richey, 1994, p. 57). Summative evaluation occurs after the institutionalization of a program because the focus is on the impact of the program (Seels & Glasgow, 2001).

Products of summative evaluation serve as information for an external audience or decision-makers to be able to analyze costs, benefits and performance improvement. The purpose is to gather information about permanent effects of a program or initiative such as effectiveness, efficiency and benefits. Summative evaluation presents decision-makers with data to assist them in determining the life of a program – whether to continue with it or redirect time, money, personnel and other resources in more productive directions (Guskey, 2000).

As with the other domains of instructional technology, work within the domain of evaluation requires careful consideration of a variety of questions. Decisions regarding evaluation are made at all levels in the process of instructional technology.

What overall process or model will be used for evaluation?

Several types of models of evaluation exist, each including the components of problem analysis, assessment for learning, formative and summative evaluation. These models assist the instructional technologist in developing a framework for how evaluation will be conducted at all stages of a project. Although a variety of types of models exist, four types stand out as extremely relevant to the work of an instructional technologist – goal-based, goal-free systems analysis and decision-making. Determining

which type of model to use involves asking why the project is being evaluated (Boulmetis & Dutwin, 2005).

Projects are evaluated for a number of reasons. If a technologist seeks to determine effectiveness of a program to funding sources or management, a goal-based model is a logical choice. Goal-free models assist technologists in learning about their programs in order to improve. If the goal is to determine the worth of a program, measure its impact or determine the effectiveness of instruction in order to promote or reduce it, decision-making or systems analysis models are most appropriate (Boulmetis & Dutwin, 2005). Specific models of evaluation are categorized according to their purposes and are described briefly below.

Tyler's Evaluation Model (goal-based), Scriven's Goal-Free Evaluation model (goal-free), Stufflebeam's CIPP Evaluation Model (decision-making) and Kirkpatrick's Evaluation Model (systems analysis) all serve as examples for the evaluation of instructional programs or interventions (Guskey, 2000).

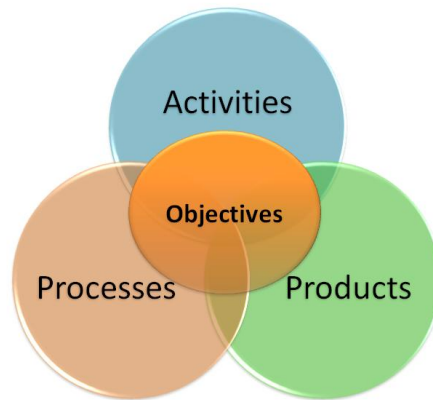


Figure 1: Adapted from Ralph Tyler's Evaluation Model (1949)

Goal-based models are typically based on Ralph Tyler's Evaluation Model (1949) in which program goals are measured against actual performance. Tyler believed that if a program is to be planned and improvement is to be made, clear goals and objectives should drive all material selection, instructional procedures and assessments (Tyler, 1949). Tyler's model includes the establishment, classification and definition of goals before any data collection is conducted. Then, data collection methods are selected by determining where and through what processes objectives would be most likely demonstrated and techniques for data collection are selected. Performance data is collected and that data is compared with the stated objectives (Guskey, 2000). While this model focuses on how performance compares with stated objectives, other approaches reject this idea, favoring the identification of needs and basing evaluation on actual effect on those needs.



Figure 2: Adapted from Scriven's Goal-Free Evaluation Model (1972)

Scriven's Goal Free Evaluation Model (1972) suggests that focusing on a program or activity's goals can be an important starting place for a technologist working in the domain of evaluation. Scriven (1972) believed that "goals of a particular program should not be taken as a given," but examined and evaluated as well (Guskey, 2000). The Goal-Free model focuses on the actual outcomes of a program or activity, rather than only those goals that are identified. This type of model allows the technologist to identify and note outcomes that may not have been identified by program designers (Guskey, 2000). Through a process of both blatant and veiled techniques, this method seeks to gather data in order to form a description of the program, identify processes accurately, and determine their importance to the program (Boulmetis & Dutwin, 2005). While this model focuses on the outcomes regardless of goals, other models focus on the processes of decision-making and providing key administrators with deep analysis to make fair and unbiased decisions.

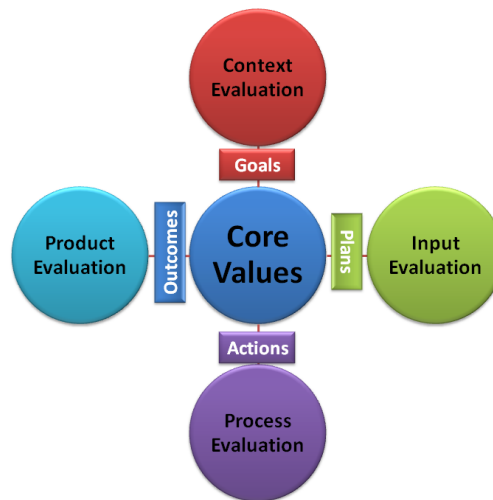


Figure 3: Adapted from Stufflebeam's CIPP Evaluation Model (1983)

The CIPP Model (Stufflebeam, 1983) focuses on the collection of four different types of data to inform the decisions of organizational administrators – context (C), input (I), process (P) and product (P). Context evaluation is similar to problem analysis in that it serves as the "planning" information, identifying problems, needs and opportunities in order to develop program objectives. Input evaluation provides information regarding the resources that are available and needed in order to achieve

identified objectives. Process evaluation evaluates whether changes are needed within an organization's work environment in order to implement, but also measures whether program elements are being implemented as intended. Product evaluation focuses on the outcomes of a program or activity and helps administrators determine whether to continue, terminate, modify or refocus (Guskey, 2000). While the ability to influence the decision-making process is relevant to the job of the instructional technologist, the ability to determine whether what he is designing, developing and implementing is working is critical.



Figure 4: Kirkpatrick's Evaluation Model (1994)

A systems analysis approach, such as Kirkpatrick's Evaluation Model is quite useful in determining the effectiveness of an intervention. Kirkpatrick (1994) identified four levels of evaluation to measure organizational effectiveness and efficiency. The levels of reaction, learning, behavior and results all contribute to the overall quality, efficiency and effectiveness of a training program or intervention (Guskey, 2000). Reaction evaluation is concerned with the feelings of participants in the program, assuming that the purpose of the training is to help them. This type of evaluation is designed to measure their satisfaction with the training they received. Learning evaluation is the process of measuring the knowledge, attitude and skills of program participants and how effective the program has accomplished the objectives. Behavior evaluation is the process of observing participants in the context of their job following training, determining how much and what type of change has taken place. Results evaluation assesses organizational change as a result of the intervention in areas such as improved productivity, better quality, reduced cost, and improved morale (Guskey, 2000).

In order to determine which evaluation model best fits a project, the technologist must answer the question, "Why am I evaluating?" (Boulmetis & Dutwin, 2005). The answer to that question will lead him to an appropriate and effective model for evaluating training programs or interventions.

What data will be analyzed during the process of needs assessment?

The work of needs assessment, although conducted within the domain of design, is actually a process of evaluation and includes the collection of data in order to make judgments regarding the design of an intervention or instructional program. Data is collected and analyzed during a needs analysis using the techniques of extant data analysis, needs assessment and subject matter analysis (Rossett, 1987). Extant data is defined as the "stuff that companies collect that represents the results of employee

performance” such as sales figures, reports, letters, attendance reports and exit interviews (Rossett, 1987, p. 25). Needs assessment data are the opinions and feelings of those who are affected by a performance problem. Subject matter analysis is conducted through the interaction with subject matter experts to derive essential information about the nature of a job, the bodies of knowledge surrounding it and the specific tasks required to perform it (Rossett, 1987). The purpose of analyzing this data is to seek optimal, actual feelings surrounding a performance problem and assisting in determining causes and developing strategies to solve it

What types of assessment tasks will be used within the learning process?

There are two main types of assessment items commonly used in an instructional program – objective assessments and performance-based assessments. Objective test items are highly structured and have a single right or best answer (multiple choice). Performance tasks are activities that learners perform in order to demonstrate their learning (i.e. write an essay, analyze a case study, utilize a scientific tool). Both types of items provide valuable information to the technologist or instructor regarding the learning process. Determining which type of task to use depends on the learning outcomes that will be measured and the data that is needed (Linn & Miller, 2005).

Objective test items are particularly efficient for measuring knowledge of facts, while performance activities measure understanding, thinking and other complex learning outcomes. In preparing the activities, performance tests require only a few tasks to be designed where a large number of questions are needed for an objective test. The amount of content that can be tested using objective activities is much higher than that of performance assessments. If control of responses is a consideration, objective items offer a much higher amount of control, since performance assessments offer the learner freedom in how to respond. Scoring and reliability are also a consideration since objective tests are scored objectively and performance tests have a degree of subjectivity, making them subject to possible inconsistent scoring. Finally, objective test items encourage students to develop knowledge of specific facts and discriminate among them in their preparation, while performance tasks encourage the organization, integration and synthesis of a large unit of subject-matter (Linn & Miller, 2005).

The main principle in assessment selection is to determine which item type that will most directly measure the intended learning outcome (Linn & Miller, 2005). The use of both objective and performance assessment activities in an instructional program will provide the technologist or instructor a comprehensive measurement of participant learning (Linn & Miller, 2005).

In the development of instructional materials, what process will be used to determine their effectiveness?

Dick, Carey and Carey (2005) advise that formative evaluation should be used at the design and development levels of the ISD process using an iterative process containing three cycles – one-to-one evaluation, small group evaluation and field trial. In one-to-one formative evaluation, gross errors such as clarity of language and accuracy of content are eliminated from instructional products. In a small group evaluation, representatives of the target population locate additional errors in the materials and management procedures. Finally, a field trial is conducted in the setting and context where instructional materials will be used to identify additional errors (Dick, Carey & Carey, 2005).

What approach will be used when conducting a summative evaluation of an instructional program or intervention?

Three aspects of projects are measured during summative evaluation: effectiveness, efficiency and benefits (Seels & Glasgow, 2001). There are a number of approaches to summative evaluation including expert judgment, operational tryout, comparative experiment, cost/benefit analysis (Seels & Glasgow, 2001).

Dick, Carey and Carey (2005) recommend the use of expert judgment and operational tryout as a combined approach to summative evaluation, identifying operational tryout as “field trial.” During the expert judgment phase, the technologist must consider whether the materials have potential for meeting the organization’s needs. The field trial, or operational tryout phase, is the process of determining whether the materials are effective with the learners in the prescribed setting (Dick, Carey & Carey, 2005). While this approach compares the performance of learners with the intended objectives, other approaches compare the performance of one group with that of the other to determine outcomes.

Comparative experiment identifies a control group in addition to the experimental group that is participating in or utilizing the prescribed intervention. This approach allows the client to observe changes under clearly defined conditions, since the control group has not benefited from the intervention. By comparing data from the two groups, judgments can be made regarding effectiveness, efficiency and the benefits of an intervention to determine whether time and resources are warranted to make the change (Seels & Glasgow, 2001). As this approach provides a clear comparison, the time and resources required to conduct such an experiment can be costly.

Conducting a cost-benefit analysis allows the technologist to analyze the costs associated with a particular intervention against the benefits that are observed. Intervention costs can be development cost, start-up cost and operating cost. Benefits can include increases in productivity, employee performance or changes in learner skills, attitude or knowledge that impact a program or organization. By analyzing these facets of an intervention, stakeholders are able to realize changes in effectiveness, efficiency and benefit.

While a number of approaches exist for summative evaluation, the product of it is a report that will inform decision-making, facilitate further intervention or identify subsequent needs. In the process of instructional technology, summative evaluation may lead to problem identification and the execution of further interventions.