

## **Domain of Evaluation**

Seels and Richey (1994) define evaluation as “the process of determining the adequacy of instruction and learning” (p. 54). They assert that evaluation occurs throughout all stages of the instructional process and affects all domains of instructional technology. Evaluation is a natural process because “we are constantly assessing the worth of activities or events according to some system of valuing” whether it is a new product or idea (Seels & Richey, 1994, p. 52).

Within the domain of evaluation, Seels and Richey (1994) makes the distinction between three levels of evaluation: program evaluation, project evaluation, and materials evaluation (p. 55). An explanation of these levels can be found in Table 3.

In order for evaluation to be successful, three areas within an instructional technology project must overlap. Instruction, objectives, and evaluation must all be interrelated so that instruction and evaluation are based on the objectives while the instruction and objectives are also evaluated (Chen, 2006).

Many forms and levels of evaluation exist including product, process, formal, informal, performance, achievement, and confirmative in addition to the sub-domains identified by Seels and Richey (1994).

The sub-domains of evaluation are:

- Problem Analysis
- Criterion Referenced Measurement
- Formative Evaluation
- Summative Evaluation

## **Problem Analysis**

According to Seels and Richey (1994), problem analysis “involves determining the nature and parameters of the problem by using information gathering and decision making strategies” (p.56). Seels & Glasgow (1998) explain that problem analysis includes needs assessment to determine gaps, performance analysis to determine if instruction is a potential solution or part of the solution, and contextual analysis is conducted to determine what learner and environment characteristics must be considered in the proposed solution.

During the early stages of the instructional design process, a need has often been identified in the given situation and a needs assessment is required in order to effectively evaluate the needs. In her book *Training Needs Assessment*, Alison Rossett (1987) describes training needs assessment as “the systematic

study of a problem or innovation , incorporating data and opinions from varied sources, in order to make effective decisions or recommendations about what should happen next” (p.3).

According to Rossett (1987), the purposes of the training needs assessment (TNA) are to gather information about optimal, actual, feelings, causes, and solutions. Optimal is what should be in the situation while actual represents what is in the given situation. Feelings are related to learner, organization, and management attitudes about the situation. Finally, causes and solutions are related to opinions and data about possible causes of the problem as well as opinions and data about possible solutions. These purposes are described in Table 4.

In order to gather information, multiple techniques can be used including extant data analysis which studies existing documentation, needs assessment which describes the opinions of stakeholders, and subject matter analysis which is related to the subject matter and included content. These techniques are described in more depth in Table 5.

Data is collected for these analyses through various tools. Interviewing, facilitating groups, and surveys are wonderful tools for collecting data on all purposes of a TNA while observation is important in collecting extant data information for optimal and actual (Rossett, 1987).

Having used Rossett’s process for planning and conducting training needs assessment, I learned that knowledge about the purposes; optimal, actual, feelings, causes, and solutions; is ultimately what makes the instruction effective because I, as the designer, am aware of all of the organization’s needs, possible causes, and possible solutions as they are designing training. Therefore, I have a clear picture of what will be needed for training and how it will be most effective.

In addition, I found from my study of Rossett’s process that I had underestimated the value of extant data in the past, but after using this technique in a large scale project, I see how valuable extant data is in identifying optimal, actual, and, at times, possible causes. In my opinion, every effort should be made by the designer to gather the necessary extant data in conducting a training needs assessment and a client who is willing to implement the resulting training program to their benefit will provide the needed data.

### **Criterion-Referenced Measurement**

Seels and Richey (1994) define criterion-referenced measurement as “techniques for determining learner mastery of pre-specified content” that “lets the students know how well they performed relative to a standard” (pp. 56-57). To measure mastery, four types of criterion referenced tests can be utilized for data collection. Entry behavior tests help ensure learners possess necessary

entry behaviors, pretests assess the learners' knowledge of the content before instruction, practice tests provide students opportunities for practice and feedback during instruction, and posttests assess students after instruction. These four test types are described further in Table 6.

In my experience in K-12 education, I found that often the entry behavior and practice tests are unutilized or misused because the focus of today's educational system is often on the pretest vs. posttest data. However, K-12 education could significantly increase their overall scores and effectiveness of the instructional aspects of education by ensuring that students have the appropriate entry behaviors for instruction and by providing a low-risk environment for practice tests. Having an environment conducive to risk-taking in learning will allow the students to practice the knowledge and skills in context and receive meaningful feedback. Therefore, they will learn from early mistakes and increase their ability levels.

### **Formative Evaluation**

According to Seels and Richey (1994), formative evaluation "involves gathering information on adequacy and using this information as a basis for further development" (p. 57). Formative evaluation usually occurs early in development and can indicate multiple revisions that should be made. For this reason, Seels and Richey (1994) explain that formative evaluation can have a wide reaching scope. In addition, process evaluation falls into formative evaluation because it is an ongoing evaluation of the processes in designing and developing instruction as well as students part in the process (Chen, 2006).

Dick, Carey, and Carey (2005) identify three phases that are essential in an effective formative evaluation. One-to-one evaluation, small group evaluation, and field trial evaluation all contribute valuable data to the formative evaluation process.

One-to-one evaluation identifies and removes the most obvious errors in instruction such as misspelled words, grammar issues, or misinformation as well as identifying early performance indications and reactions to the material by the learners (Dick, Carey, & Carey, 2005).

Small group evaluation first evaluates the effectiveness of the changes made after the one-to-one evaluation and evaluates "whether the learners can use the instruction" in the intended format (Dick, Carey, & Carey, 2005, p. 288).

Finally, Dick, Carey, and Carey (2005) describe field trials are conducted in an attempt to use a learning situation as similar as possible to the end performance context as well as if prior changes made during both one-to-one and small group trials were effective.

In the multiple design and development projects I have been part of throughout the MIT program, I have found the formative evaluation process to be an invaluable practice that allows the designer to ensure the effectiveness of their program or package as well as its accuracy and professional appearance, before distributing the package for implementation. Often small errors with usability including slight omissions or technology glitches in the instructional package can be corrected and worked through before releasing the materials for widespread use. This process often prevents large scale trouble shooting efforts later in the process thus, reducing expense and time needed for implementation.

## **Summative Evaluation**

Seels and Richey (1994) define summative evaluation as the “gathering of information on adequacy and using this information to make decisions about utilization” (p.57). Because of the revisions made after formative evaluation, summative evaluation is much narrower in scope and is concerned with the effectiveness of the total product rather than specific revisions to be made (Seels & Richey, 1994). Therefore product evaluation is also a type of summative evaluation.

Within summative evaluation there are two phases: expert judgment and field trial. Expert judgment allows the evaluators to gather information about the instruction and its potential for meeting the organizations needs while field trial evaluates the effectiveness of instruction. These phases are explained further in Table 7.

One model I have used for summative evaluation is Donald Kirkpatrick’s Four Levels of Evaluation (1979). This model is helpful, in my opinion, because so much information is gathered about the program or package being evaluated. As you can see in Figure 13, this model addresses four levels of information needed to conduct a summative evaluation beginning with the most basic foundational level, reactions, and ending with the most specific abstract level, results. More extensive data is collected on the learners’ most basic needs and reactions because these in turn affect the other areas of evaluation. The pyramid structure is an important visual tool for me because it represents the different levels of evaluation and how they build on one another to provide both detailed and large-scale feedback about the program or package.

Reactions are defined as “how well the participant liked a particular training program” (Boverie, Sanchez-Mulcahy, and Zondlo, 1994, p. 4). Reactions may include if the learner liked the instructor, how they felt about the room, or if they liked the topic, in addition to many other topics. On the other hand, learning is defined by Boverie, Sanchez-Mulcahy, and Zondlo (1994) as “principles, facts, and techniques that were understood and absorbed by the participants (p. 4). This includes knowledge and skills as well as attitudes that the learner gained or developed as a result of the instruction.

Transfer of learning assesses the “transfer of training skills or knowledge to the job” (Boverie, Sanchez-Mulcahy, and Zondlo, 1994, p. 5). This will include how the learner performs the after instruction compared with how they did prior to instruction. Finally, results measure the impact on the organization (Boverie, Sanchez-Mulcahy, and Zondlo, 1994, p. 8). These results are the hardest category to evaluate because very few quantitative measures exist that effectively measure outcomes. Instead, “anecdotal efforts to measure results” must be used (Boverie, Sanchez-Mulcahy, and Zondlo, 1994, p. 8).

In his article *Measuring ROI: The fifth level of evaluation*, Jack Phillips proposes that a fifth level of evaluation exists entitled Return on Investment (ROI). ROI measures the “annual net program benefits divided by program costs, where the net benefits are the monetary value of the benefits minus the costs of the program” (p.12). In simpler terms the ROI (represented by a percentage) is equal to the benefits minus the cost multiplied by one hundred and divided by the costs. This formula is shown in Figure 14.

The basic question being asked by the ROI model is “did the monetary value of the results exceed the cost for the program?” (Phillips, 1996, p. 12)

## **Conclusion**

Evaluation as a domain of instructional technology provides the designer with information regarding the effectiveness of the design and development of their project, information about whether the utilization of the product is correct and effective, and provides formative revisions as well as overall data on effectiveness. I believe the four sub-domains of evaluation serve to cross the domains and aid in collection of evaluation data at all levels and stages of the project. Data collected for evaluation may be qualitative for objective measures or quantitative for more subjective information needs (Seels & Richey, 1994). In my opinion, whatever the data that is collected during evaluation and no matter which models are used, the goal should be to improve the instruction in order to make the product as accurate and effective as possible in serving its purpose.

In addition to the four sub-domains outlined by Seels and Richey, other types of evaluation are being identified that provide different information. One important type of evaluation not defined by Seels and Richey is confirmative evaluation. Confirmative evaluation is one aspect of James D. Russell’s Evaluation Model which intertwined formative and summative with confirmative. While formative evaluation is conducted during the process and summative is an evaluation of the effectiveness of the entire package, confirmative evaluation serves to confirm the long-term success of change by looking at knowledge and skill retention (Chen, 2006).