

The Domain of Evaluation

Seels and Richey(1994) define evaluation domain as “process of determining the adequacy of instruction and learning”(Seels & Richey, 1994, p. 54). The evaluation domain is guided by the systematic acquisition and assessment of information to provide useful feedback about products or learning processes. Most often, feedback is perceived as "useful" if it aids in decision-making. But the relationship between an evaluation and its impact is not a simple one. Despite this complexity, the major goal of evaluation should be to influence decision-making or policy formulation through the provision of empirically-driven feedback.

There are many different types of evaluations depending on the object being evaluated and the purpose of the evaluation. For example, the object of evaluation may be a program, a project, or a product. In addition, the evaluation may focus on determining the merit, worth or value of a program (summative) or it may emphasize strengthening or improving a program or a product (formative). Thus, the most important basic distinction in evaluation types is that between **formative** and **summative** evaluation.

Evaluation process often begins with needs analysis, and clarification of goals and constraints. The instructional designer then collects data regarding the program, process, or product. This data can be collected using observations, surveys, questionnaires, or testing. Data is then analyzed and compared to the instructional goal to determine the overall effectiveness. This process is cyclical once revisions have been made.

An instructional designer uses evaluation throughout the Instructional Systems Design (ISD) process. In the beginning of a project an instructional designer conducts a needs analysis to determine the need or gap. Once a learning solution (a program or a product) is designed and developed or in the process of design and development, a series of formative evaluations are conducted to ensure the efficiency and effectiveness of courses or materials being developed. The instructional designer often uses criterion-reference tests to determine whether the learner has met the performance objectives or whether or not the training has been effective. After the program has been implemented, the instructional designer determines the goal of the program and may conduct a summative evaluation to determine the overall effectiveness.

There are four sub-domains within the domain of evaluation: problem analysis, criterion-referenced measurement, formative, and summative evaluation.

Problem Analysis

Problem analysis involves determining the nature and parameters of the problem by using information-gathering and decision-making strategies (Seels & Richey, 1994, p. 56). Problem analysis is also known as Front-End Analysis or Needs Assessment. This is a systematic method of determining the desired or optimal status and the actual status to establish the gap or need. This enables the instructional designer to verify the problem and identify solutions. By determining the gap, one is able to define the goals and

constraints during problem analysis. An instructional designer uses problem analysis as the program is being conceptualized. Information is gathered using a variety of sources such as interviews, surveys, observations, and extant data analysis. This analysis typically includes learner analysis, environmental analysis, and contextual analysis (Dick, Carey, Carey, 2005).

Allison Rossett's Training Needs Assessment offers techniques on gathering data to perform a problem analysis. Training Needs Assessment focuses on the "actuals" and the "optimals" by analyzing the skills, knowledge, and attitudes to determine a "gap" or need. Other methods include Dick and Carey's (2005) Front-End Analysis which includes performance analysis, needs assessment, and sometimes job analysis.

Criterion-Referenced measurement

Criterion-referenced tests are also referred to as content-referenced, objective-referenced, or domain-referenced. In other words, a specific list of objectives or a list of identified content are used to construct measurement tools that can document whether the learner has gained the knowledge, skills, or attitude laid out by the performance objective. Criterion-referenced measurements provide information about a person's mastery of knowledge relative to the objectives they were to accomplish. This demonstrates to the trainer and the learner whether the objectives have been met. The learner is not ranked against others as in norm-referenced measurement but is only compared with himself/herself.

Criterion-referenced measurement can be used for pre-tests to determine whether the learner has the pre-requisite skills necessary for the training. It can also be used for post-tests to determine whether one has mastered the given objectives. Instructional designers typically use criterion-referenced tests since the outcome typically focuses on the given goal and objectives laid out by the training or program.

Formative Evaluation

Formative evaluation is conducted throughout the development stage in order to improve or refine the effectiveness of a program or training. This type of evaluation is typically done by an internal evaluator to increase the effectiveness and efficiency of the program. This evaluation focuses on both the collection of qualitative and quantitative data in order to revise instruction. Formative evaluation is most valuable when conducted during development and the tryout phases.

An instructional designer often uses formative evaluation throughout the instructional systems design process. While materials are being developed, an instructional designer conducts formative evaluation to determine the effectiveness of the materials. It is important to complete this early in the process to ensure working material when it is time to implement the project. Prior to implementation, the instructional designer conducts another formative evaluation.

There are three stages of formative evaluation: one-to-one, small group, and field trial. The learners who are selected for formative evaluation should be a representative of the target population. During

the one-to-one phase, the instructional designer works with individual learners to modify the most obvious errors in the instruction and to obtain initial reactions to the content by the learners. The small group usually consists of eight to twenty learners. The purpose of the small group evaluation is to determine the effectiveness of the changes made after the one-to-one evaluation and to identify any learning problems that the learners may have. During the third stage of evaluation, field trial, the emphasis is on the procedural methods and is as close to the “real world” as possible. This is done to ascertain whether the changes made after the small group evaluation were effective. Also, data is gathered to determine whether the instruction can be used in the context in which it was intended by evaluating the learners’ reactions to the material and process along with evaluating the reliability of the material and process. The evaluation can be done through a variety of methods such as observation, debriefing, or short tests (Dick, Carey and Carey, 2005).

Formative evaluation can be completed on a micro or macro level. On a micro level the instructional designer is concerned with whether the learner acquired specific skills or knowledge. At the macro level, the instructional designer looks at the implementation of the entire system.

Summative Evaluation

Summative evaluation judges a programs merit or worth one it has been implemented for the targeted learners. This evaluation is usually conducted by an external evaluator for the benefit of an external audience or decision maker. With the completion of a project, the instructional designer analyzes how the group did with regards to the leaning task. By evaluating the group, the instructional designer can evaluate the learning materials and the learning process.

Summative evaluation has two phases: expert judgment and field trial. The expert judgment phase examines congruence analysis, content analysis, design analysis, and feasibility analysis. Congruence analysis aligns the program with the organizational needs and the resources the organization has available for purchase and use. Content analysis identifies the goals and sub goals. This framework is then compared with the materials to review the quality. Design analysis is used to evaluate the adequacy of the components of the instructional strategies used in the materials. Feasibility analysis relates to the utility of the material. This includes availability and durability of materials or other special factors such as equipment or leaning environments. Field trial looks at outcome analysis, impact on learners, impact on the job, impact on the organization, and management analysis (Dick, Carey, & Carey, 2005).

Many models are used for evaluations. Some of the most popular are Kirkpatrick’s Four Levels of Training, CIPP, and Connoisseurship Model on evaluation.

There are many evaluation models an instructional designer can use to guide the process of evaluation. The Goal Free evaluation is designed to find out what a program is actually doing without knowing the purpose of the program. The evaluator will need two types of information. First, they will need to identify the outcomes resulting from the project and then construct a profile of the needs from the

target population (Manfredi, 2003). If there is a positive impact on the needs of the target audience then the program will be positively evaluated.

The Connoisseurship Model is subjective in nature. The evaluator serves as a critic with the program under review. The evaluator serves as a participant observer and renders judgments based on personal expertise. Two aspects of this model are critical. Members must have confidence to submit to the uncertainty of this type of evaluation. Also the evaluator's credibility and expertise are essential. This type of evaluation is useful when administrators would like to use language and publicity to promote the value of the program (Robbins, 2006).

The Kirkpatrick model is based on four levels: Reaction, Learning, Performance, and Impact. Reaction measure the learner's perception of the course. Level two addresses the extent to which the learners had a change in attitude, or increase of knowledge or skill. Behavior is measured in level three. This primarily addresses whether the student is now performing the new skill or knowledge in the workplace. Level four addresses the overall effectiveness of the program or training.

CIPP has been described as a roadmap for evaluation. CIPP stands for Context, Input, Process, and Product. Context evaluation determines the identity of the target population by defining the parameters of the organization. Information can be gathered through interviews or surveys depending on the size of the organization. Assessing financial and opportunity costs along with other potential problems with regard to successful implementation occur during input evaluation. Once the results have been identified, the process evaluation deals with the possible need to restructure. Product evaluation assesses the effects of the program. Advantages to using the CIPP model include comprehensiveness, flexibility, integration and decision-orientation (Stufflebeam, 2002).

Confirmative Evaluation

Increasingly, instructional designers are conducting confirmative evaluations. Confirmative evaluations are similar to formative and summative in respects to gathering information. Confirmative evaluation relies on multiple data-gathering instruments such as questionnaires, interviews, performance assessments, self-reports and knowledge tests. Confirmative evaluations occur after the implementation has occurred and been in use for a reasonable amount of time. This type of evaluation assesses not only the competence of the learners but also the continuing effectiveness of a program. The challenge for instructional designer is to continuously collect data in regards to a program.

For example, a graduate program may want to assess the effectiveness of their program 5 years after students have graduated. The instructional designer will have to collect data from the learners in regards to whether or not they are still competent with the information and whether they are utilizing information or knowledge gained in the program (Schenkman, 2004).