Definition of Evaluation

"Evaluation is the process of determining the adequacy of instruction and learning" (Seels & Richey, 1994, p. 54). The systematic process of evaluation allows instructional designers to establish the gap(s) in performance and/or knowledge (needs assessment), develop and administer quality tests which accurately measure the performance objectives (criterion-referenced measurement), identify and apply necessary changes to the project while still in the development stage (formative evaluation), and to determine the relative success of a project after full implementation and institutionalization (summative evaluation).

Needs Assessment

Needs assessment or front-end analysis is: "the systematic effort that we make to gather opinions and ideas from a variety of sources on performance problems or new systems and technologies" (Rossett, 1987, p. 62). For an instructional designer this is often the first step in the instructional design process. Rossett (1987) has identified five items instructional designers look for when conducting a needs assessment:

- 1. optimal performance or knowledge
- 2. actual or current performance or knowledge
- 3. feelings of trainees and significant others
- 4. causes of the problem from many perspectives
- 5. solutions of the problem from many perspectives

Data is gathered from various sources using surveys, interviews, and observations first to identify any performance or knowledge-based problems and next to recognize any gaps between optimal and actual performance. The instructional designer then must prioritize the gaps or needs between the actual and optimal performance to identify needs (Seels & Glasgow, 1998). There a number of different models that can be used to perform a front-end analysis including Allison Rossett's Needs Assessment Model (1987) and Roger Kaufman's Model (1993).

Criterion-Referenced Measurement

Tests can be divided into two categories: norm-referenced tests and criterion-referenced tests. Norm-referenced tests are designed to classify students across a continuum of achievement from high achievers to low achievers but not to measure knowledge (Bond, 1996). With this type of test, the assessment is initially administered to a representative group of the target population. This original group's scores set the "norm" to which the general population will be compared.

Criterion-referenced tests (CRT), on the other hand, "report how well [learners] are doing relative to a pre-determined performance level on a specified set of goals or outcomes" (Bond, 1996, p. 2). Criterion-referenced measurement allows the instructional designer to accurately and effectively measure the learner's acquisition of knowledge, skills, or attitudes. CRTs should directly align with the performance objectives created in the analysis stage.

Instructional designers use criterion-referenced tests exclusively over norm-referenced tests because CRTs accurately measure knowledge acquisition instead of achievement differences among students.

Formative Evaluation

The purpose of formative evaluation is to "locate weaknesses and problems in the instruction in order to revise it" (Dick, Cary & Carey, 2001, p. 359). Formative evaluation allows the instructional designer to correct any mistakes before the product has been implemented, saving the project numerous resource costs and keeping it on time, under budget and to specifications. Subject-matter experts play a key role in the process of formative evaluation. They lend their expertise to ensure the design and products match the stated objectives.

There are three stages of formative evaluation: one-to-one, small group, and field trial. During the one-to-one or clinical evaluation stage, the instructional designer works with individually with learners to acquire data to revise the instruction (Dick, Carey, & Carey, 2001). For the second stage, a group of between eight and twenty learners from the target population, study the materials and are tested on the material to obtain the necessary formative evaluation data. The final stage, entitled field trial, tests learners in an environment and context as close to the "real world" as possible. The number of learners in this last stage varies but most often thirty participants is adequate (Dick, Carey, & Carey, 2005).

The formative evaluation process is often conducted "in-house" by a member of the design and development team. This "internal evaluator" often has a personal investment in the material and thus seeks accurate judgments regarding the materials in order to produce the best materials possible (Dick, Carey, & Carey, 2001).

The instructional designer revises the materials using the data from all three stages. Ideally major problems are discovered in the formative evaluation process because once the project has reached the production stage it is very costly and time prohibitive to make changes.

Summative Evaluation

Summative evaluation is meant to measure the long-term effects of the instruction and whether the initial knowledge or performance problem, identified in the needs analysis, has been solved. There are two stages of summative evaluation: expert judgment and

field trial. The expert judgment sub-stage determines whether the materials have the potential to meet the organization's defined instructional needs (340). The field trial sub-stage involves the actual implementation of the instruction within the organization using the target learners.

After both an expert judgment and field trial have been conducted the resulting data is used by the organization's decision makers to decide whether the current materials should be maintained or new materials adopted (349).

An third party (external evaluator) is hired to conduct the summative evaluation as it can be difficult for a member of a project's design and development team to remain objective when comparing their product to another.

One summative evaluation model used extensively in the filed to evaluate training is Donald Kirkpatrick's Four Levels of Evaluation (Kirkpatrick, 1998). This model's four levels consist of: reaction, learning, transfer, and results. Each level is built on the information gathered in the previous level and so the most time-consuming but telling evaluation occurs at the fourth (results) level.

Kirkpatrick Four Levels of Evaluation Model, (Kirkpatrick, 1998)



Figure 1: Four Levels of Evaluation Image adopted from: coe.sdsu.edu/eet/Articles/ k4levels/start.htm

The first level of evaluation is "Reactions". Reactions are gathered immediately after an instruction is delivered. This information is often gathered using a tool commonly referred to as "smile sheets." Smile sheets measure the learner's feelings towards the facilitator, the instruction, and the environment. This first level does not measure learning, but if the learner has negative feelings towards the training material, facilitator, or the learning environment it will be difficult for learning to occur.

The second level, entitled learning, measures the learner's advancement of a skill, knowledge, or attitude. A proctored individual pre- and post-test are most often used in this level but variations include a self-test or team-test.

Transfer, Kirkpatrick's third level evaluation, is meant to measure if the learner applies the new knowledge, skills, or attitude, acquired from the training, in their job on a regular

basis. This level is difficult to measure as it is hard to predict when the change in behavior will occur. Many instructional designers see this level as the most important level of evaluation.

The fourth level, results, is an attempt to measure the training's return on investment. Was the training responsible for increased production, improved quality, decreased costs, reduced accidents, increased sales, and/or higher profits? This final level is used to justify the cost of training to managers and executives.

In addition to Kirkpatrick's Four Levels of Evaluation model, there are a number of other models used by instructional designers including: The Flashlight Triad Model and Daniel Stufflebeam's CIPP Evaluation Model.

The domain of evaluation is a key component in any instructional technology project. The four sub-domains (front-end analysis, criterion referenced testing, formative evaluation, summative evaluation) help to guide the practice of evaluation.