How to Evaluate a Worker T&D System: Evaluative Criteria

Moses Rivera, Ph.D.

August 4, 2025

Downloads

- PDF version of these lecture notes.

Beyond the ***Lecture Notes***: Insights and Elaborations

- In this video, I don't summarize these lecture notes. Rather, I provide related insights and elaborations.

[video]

Introduction

- When designing a T&D system, the best first step is to start with the end in mind. What **outcomes** are desired from this T&D system? And how will those outcomes be **measured and evaluated**? That is the focus of this lecture.

What criteria will we use to evaluate a T&D system?

- The following is a simple yet comprehensive set of evaluative criteria categories:
  1. Persons' **attitudes and other beliefs** about the T&D system.
  2. Evidence collected during the training session(s), about whether/how much the trainees are **learning**.
  3. Evidence of **ToT** (**transfer-of-training**), especially long after the training session(s).
  4. **Costs** of training.
  5. **Other outcomes** that may be attributable to the T&D system.
- Below are further details about each of the above five categories.

Persons' attitudes and other beliefs about the T&D system

- Recall that an **attitude** is:

"a relatively enduring and general evaluation of an object, person, group, issue, or concept on a dimension ranging from negative to positive." (online APA Dictionary of Psychology, 2025)

- In other words, attitudes about the T&D system are *general beliefs* about how *good/bad* the T&D system is overall, or about some specific element of the T&D system.

- Other beliefs (about the T&D system) may provide further specific descriptions (from the perspective of whoever has the belief), and they may focus on evaluative questions, such as these:
  - was it well-organized?
  - was it relevant?
  - was it fun and effective?
  - was it worth my time, money, etc.?
- It's important to measure the attitudes and other beliefs from:
  - the **trainees**;
  - the **trainers**;
  - **other persons involved** (e.g., stakeholders).
- Attitudes and other beliefs can be measured at multiple time-points:
  - **long before** the T&D system is implemented;
  - **immediately before** a T&D session;
  - **during** a T&D session;
  - **immediately after** a T&D session;
  - **long after** a T&D session.
- Evidence about attitudes and beliefs can come in two broad categories. You can collect:
  - **self-report data** (e.g., the people who are reacting will tell you about their reactions);
  - **anything else that's not self-report**. Sometimes people use the word ***other-report*** to refer to data that was provided by someone who is describing another person (e.g., a coworker or supervisor describing the focal worker).
    * This also includes data that was automatically collected from a sensor, computer, etc. However, remember the assumptions you are making when you make an inference about someone's attitudes/beliefs, based on the data you're using.
  - You can use any of the typical ways that researchers use to collect data (e.g., questionnaires, interviews, observations, using archival data, etc.)

During the training session: Evidence of learning

- The online APA Dictionary of Psychology (2025) defines ***learning*** as:

  > "the acquisition of novel information, behaviors, or abilities after practice, observation, or other experiences, as evidenced by change in behavior, knowledge, or brain function."

- Thus, if we want to be sure that a person has learned, we need to see **evidence of change** in the person.
- Evidence of learning can be collected in many ways. Two broad categories of assessments are:
  - **formative assessments**: also known as *feedback evaluation*, this is a low-stakes assessment, typically during a trainee's progression in some training module (rather than at the conclusion). In addition to measuring the trainee, its primary goal is to give feedback to the trainee about their performance on what is being assessed.
  - **summative assessments**: also known as *terminal assessment*, this is an assessment that is conducted at the conclusion of the trainee's progress in some training module. Its goal is primarily to measure the trainee, rather than to give them feedback.
- Regardless of whether you're using formative or summative assessments, you can structure the assessments in all the typical ways that researchers use to collect data (e.g., questionnaires, interviews, observations, using archival data, etc.)

After the training session: Evidence of ToT

- **ToT** (**transfer-of-training**) is the process in which the worker implements into their work (or real-world context) what they were supposed to learn from the training.
- It is useful to focus on at least two *dimensions* of ToT:
  1. **Generalization**: the extent to which the changes in the trainee are applied to different contexts (e.g., places, situations, etc.) instead of only contexts that are similar to the training context.
  2. **Duration across time**: the extent to which the changes in the trainee persist across time.
- It's important to collect evidence of ToT at multiple time-points after the training session(s). Depending on the desired results, you could measure ToT:
  - immediately after;
  - a few weeks after;

  - several months after;
  - more than a year after the training session(s).
- To measure ToT, you can use any of the typical ways that researchers use to collect data (e.g., questionnaires, interviews, observations, using archival data, etc.). Here are some ideas:
  - **knowledge-checks** (e.g., written questions, interview);
  - **observations** from coworkers and/or supervisors;
  - **automatically-generated data** (e.g., from sensors or computers) about trainee's behaviors, performance, etc.

Costs of the T&D system

- *Costs* of the T&D system can refer to any resource that is used or consumed because of the T&D system. Typical examples include:
  - **Money** (e.g., to purchase equipment and supplies; rent a venue; pay the persons who built/delivered the T&D system; pay for food). Some of the costs might be **upfront**, whereas others can be permanently **ongoing**.
  - **Person-hours**:
    * a person-hour is a unit of measure that represents the amount of work the person can complete in one hour. For example, if you pay three sales workers who are each working 4 hours today, then you are paying for $3 \times 4 = 12$ person-hours on sales workers. Not all person-hours are equivalent in terms of money nor actual work outputs. For example, supervisor pay tends to be higher than entry-level pay, therefore supervisor person-hours tend to be more expensive.
    * A T&D system can cost person-hours because of several reasons, including:
      · **time during training** spent away from the worker's typical tasks;
      · **additional ongoing person-hours** that are needed to support the T&D system (e.g., permanent changes in administrative procedures)
  - **Other consumable resources** that the organization already owns.
- One type of cost–benefit analysis is called a **utility analysis**, which is a technique to **quantify the value of a training program in monetary terms**. It seeks to estimate how much the performance improvement from training is monetarily worth to the organization. In a typical utility analysis, you would do this by measuring such things as:
  - the effect of the T&D system on the workers' job performance;
  - the increase (or decrease) in revenue and/or profit as a result of the change in workers' performance;

- the number of workers who were trained;
- the expected (or actual) duration of the aforementioned changes;
- the costs associated with the T&D system.

Other outcomes that may be attributable to the T&D system

- Besides the aforementioned outcomes (e.g., changes in attitudes, beliefs, knowledge, skills, other abilities, etc.), we can also measure other outcomes that are typically important in a work context. Some of them may describe a single person and/or a team and/or the entire organization. Typical examples include:
  - Reduced **turnover** (i.e., workers leaving the organization);
  - Enhanced **quality** of the organization's final products/services;
  - Fewer production/service **mistakes**;
  - Fewer **injuries**;
  - Increased **revenue**, **profit**, **ROI** (return on investment), etc.
- You can use any of the typical ways that researchers use to collect data on these outcomes (e.g., questionnaires, interviews, observations, using archival data, etc.)

Historical models for evaluating a T&D system

- You will very likely encounter people referring to the **Kirkpatrick model** for training-evaluation. Donald L. Kirkpatrick, Ph.D., developed in the 1950s–1960s what eventually became known as the *Kirkpatrick model* for evaluating a training procedure. Today, it is good to be familiar with it because so many people still use it and refer to it, but modern sets of evaluative criteria and techniques are more comprehensive. He focused on four types of criteria, which he labeled as such:

  1. **Reactions** (How did participants feel about the training?)
  2. **Learning** (What did participants learn?)
  3. **Behavior** (Has there been ToT?)
  4. **Results** (Has there been impact on organizational outcomes?)

- There are many other similar sets of criteria categories—many of which built upon Kirkpatrick's—but that's beyond the scope of this course. You can learn about those easily via an internet search.

Suggested Readings

Items indicated with an asterisk (*) are available in the Zotero Group Library for this course.

- *Training Evaluation. (2023). In R. A. Noe, *Employee Training and Development* (Ninth edition, pp. 243–288). McGrawHill Education.

    - Our Zotero Group Library only includes chapter 6 (*Training Evaluation*) of Noe's (2023) book, in compliance with U.S. copyright regulations on fair use for educational purposes.
- *Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's Levels of Training Criteria: Thirty Years Later. *Personnel Psychology*, *42*(2), 331–342. https://doi.org/10.1111/j.1744-6570.1989.tb00661.x
    - A scholarly critique of Kirkpatrick's evaluation "model".