Learning Evaluation Theory:

Kirkpatrick Model

Provided by WatershedLRS.com

Introduction

It's always been important to evaluate learning programs. Without evaluation, how do we know if the courses, resources, and activities we lead people through are having a positive impact or just boring our learners to death? We know from studies that learning programs can have a dramatic impact on business performance, but at the same time we've all experienced courses that were a waste of time.

Evaluation tells us which of your learning solutions actually work and why. It's the first step in making sure more of our future learning solutions work, too. Learning evaluation certainly isn't new, but it is an area that is generally not done well in Learning and Development (L&D). There are more learning evaluation models and theories than you can count, but they are rarely implemented well in practice.

Recently, L&D leaders are reporting more demand from their organizations to demonstrate the effec-

tiveness of learning programs with data. At the same time, technologies are becoming available to enable L&D to do just that. It's learning's turn to prove it's worth to the business with evidence-backed facts and figures. Learning has the potential to be really transformative for the business, especially when programs are based on evidence-driven decisions. We've got an exciting time ahead!

This whitepaper explores the Kirkpatrick Model— one of many evaluation models—and its usefulness to L&D departments.

Kirkpatrick's Four Levels of Evaluation is the most ubiquitous and well known learning evaluation model. Defined back in 1959, it's had plenty of supporters, critics, and challengers over the years. We'll look at alternative models in subsequent whitepapers.

An Overview of Kirkpatrick's Four Levels of Evaluation

1 Reaction

What did learners feel about the learning experience? Was it enjoyable? Did they like the trainer?

This level is normally captured by surveys following the training.

2 Learning

Did learners actually learn anything? Did their knowledge and skills improve?

The level is normally captured by assessments at the end of the training, and sometimes at the start to illustrate a difference. With a lot of e-learning content, Level 2 is the only level that's measured.

3 Behavior

Did learners actually do anything different as a result of the training? For example, if training was designed to encourage salespeople to discuss customers' problems before proposing solutions, are the salespeople who completed the training following through?

This level is sometimes evaluated by surveying learners and/or their manager sometime after the training. Often it is not measured at all.

4 Results

What was the effect of the training on the business as a whole? For example, has there been an increase in sales?

This level can only really be measured by looking at business data relating to the training. Typically, this data is captured by the business, but it's often not compared to training data. Furthermore, L&D departments may not have access to it. Where data is captured, the challenge at this level is demonstrating the impact of the learning experience among the many other factors that can affect the metric.

Kirkpatrick & xAPI (Tin Can)

A common criticism of Kirkpatrick isn't the model itself, but how it's applied in practice. Organizations generally have processes for evaluating at Levels 1 and 2, but then either don't get around to or aren't able to evaluate Levels 3 and 4.

The Experience API (a.k.a., Tin Can or xAPI) interoperability specification makes it easier to evaluate at all four levels, especially Levels 3 and 4. xAPI records learner behavior either by integrating xAPI directly into business systems to record learner activity, or by providing mechanisms for learners to record and reflect on their performances.

Some organizations, for example, give learners mobile apps to photograph or video their work to be assessed by supervisors or mentors. These assessments can then be compared to data from the learning experiences themselves to measure the effectiveness of the experiences. Business systems also contain data about the impact on the business, which xAPI can pull into a Learning Record Store (LRS) alongside data about the learning experience and other evaluation data.[1]

Just Level 4?

Another criticism of Kirkpatrick is that Levels 1 to 3 are simply irrelevant. Investments in learning and development are (or should be) intended to drive positive business results. Therefore, the impact on business key performance indicators (KPIs) is all that needs to be measured. If the business goal is achieved, why does it matter what employees learned or how their behavior changed? Presumably, employees learned and did what we wanted them to, right?

Perhaps. Or maybe something else led to the observed business results. Without the data from Levels 1 to 3, there's no way to tell the whole story and fully understand how the end result was achieved.

Data from Levels 1 to 3 are especially important when the desired business result isn't achieved because they help pinpoint and analyze the training elements need to be changed. Perhaps the training successfully changed behavior and resulted in the sales team focusing on their customers'

challenges, but it didn't result in increased sales. This finding challenges the assumption that focusing on customers' problems—instead of providing our solutions—was a desired behavior.

Based solely on Level 4 data, we might have assumed that the training had failed to change behavior. While, in reality, the training worked, but the behavior didn't work.

> **Investments in L&D are intended to drive positive business results.**

Is Level 1 Meaningless?

Some critics have cited evidence that there's little to no correlation between what learners think of the learning experience (Level 1) and their learning or behaviors (Levels 2 and 3). For instance, learners may dislike the learning experience, but still benefit from it. Or, they may love the time away from their normal work routines, but learn nothing from the experience. This is an important criticism that underlines the point that Level 1 evaluation on its own isn't effective.

While Level 1 evaluation is less important than the higher levels, it does have one important advantage over those levels: timeliness. You can respond to data about a poor learning experience right away, whereas business impact from a learning intervention will take some time to manifest. Even behavior and learning are difficult to measure immediately, especially if we want to ensure that the learning has stuck (remember the forgetting curve).

Getting immediate feedback from the learner right after the experience (or even during) is the best way to quickly identify challenges with the learning solution—such as a broken website, a trainer who failed to arrive, or any other obstacle that stands in the way of a learning solution having its intended impact. It could also be something more minor, such as a technical problem with a specific learning interaction or an ineffective trainer.

The key here is to collect, review, and act on the feedback as fast as possible. This means monitoring learner feedback as it comes in and providing channels for learners to communicate with you at any time, not just at the end of a particular learning experience.

Level zero

The Kirkpatrick Model assumes that we've already implemented a learning solution and now want to know if it's effective. Yet, most learning (perhaps as much as 90 percent) doesn't happen during formal learning solutions. That's why many organizations are eager to first discover the types of learning that are happening within their organizations and then evaluate the resulting impacts. We could consider discovering "what learning experiences are occurring" as a level below "what learners feel about the experience" (i.e., level zero).

It's important to understand these informal learning events and their impacts on behavior and business results. After all, if we don't know these events are happening, then we can't influence them. Although we can't force workplace and social learning, we can shape, encourage, and discourage learning (i.e., not all learning is positive; bad habits are learned). The first step is understanding these learning experiences.

Kirkpatrick forgets remembering

Another important aspect missing from Kirkpatrick is remembering learning and persistence in behavioral change. The model doesn't say anything about the ongoing measurement of the four levels over time. Perhaps the training does have an initial impact, but then the learning is forgotten and the impact fades[2]. Learning solutions that include reminders are generally more effective and it's important that evaluations are ongoing and measure the lasting impact.

Kirkpatrick Summarized

Kirkpatrick identified four levels of evaluation:

reaction | learning | behavior | results

Level 4 is critical, but all levels are important to identify what did and didn't work.

We can use xAPI (a.k.a., Tin Can) to collect data at all four levels.

Learners' reactions are poor indicators of expected business results, but provide early warnings for surface-level problems.

Most learning experiences are informal and work based, but are often unevaluated because organizations don't know about these experiences.

Kirkpatrick forgets remembering; we need to measure the long-term impacts of our learning solutions. Otherwise, we risk applauding our learning experiences as successes when, in fact, learners may have forgotten everything we've taught them by the time they need it.

Upcoming Publications

This is the first in a series of whitepapers covering learning evaluation theory. Subsequent installments in this series will explore other models of learning evaluation emerging technologies that can support evaluation of learning, and a seven-step evaluation process that you can implement in your organization today.

About the Author

Andrew Downes
Learning & Interoperability,
Watershed

Andrew Downes has a background in instructional design and development, creating learning content, experiences and platforms in both corporate and academic worlds.

Now a consultant with Watershed, Andrew is a leading thinker in Evidence Driven Learning Strategy and Learning Systems Interoperability. Andrew helps organizations to develop learning strategies based on evidence and evaluation, working with vendors to get at data from the products those organizations use.

One of the authors of the Tin Can/Experience API (xAPI) specification and much of the content on tincanapi.com, Andrew has delivered numerous presentations, webinars and training sessions internationally. Andrew is well recognized as a reliable expert in the xAPI world.

References:

Kirkpatrick, Donald L. *Evaluating Training Programs: The Four Levels.* San Francisco, CA: Berrett-Koehler, 1998. Print.

[1] Downes, Andrew. *Kirkpatrick's Training Evaluation Model.* ExperienceAPI.com. Rustici Software, n.d. Web. <http://experienceapi.com/kirkpatrick/>.

[2] University of Waterloo. "Curve of Forgetting." Counselling Services. N.p., 07 Mar. 2012. Web.

Watershed