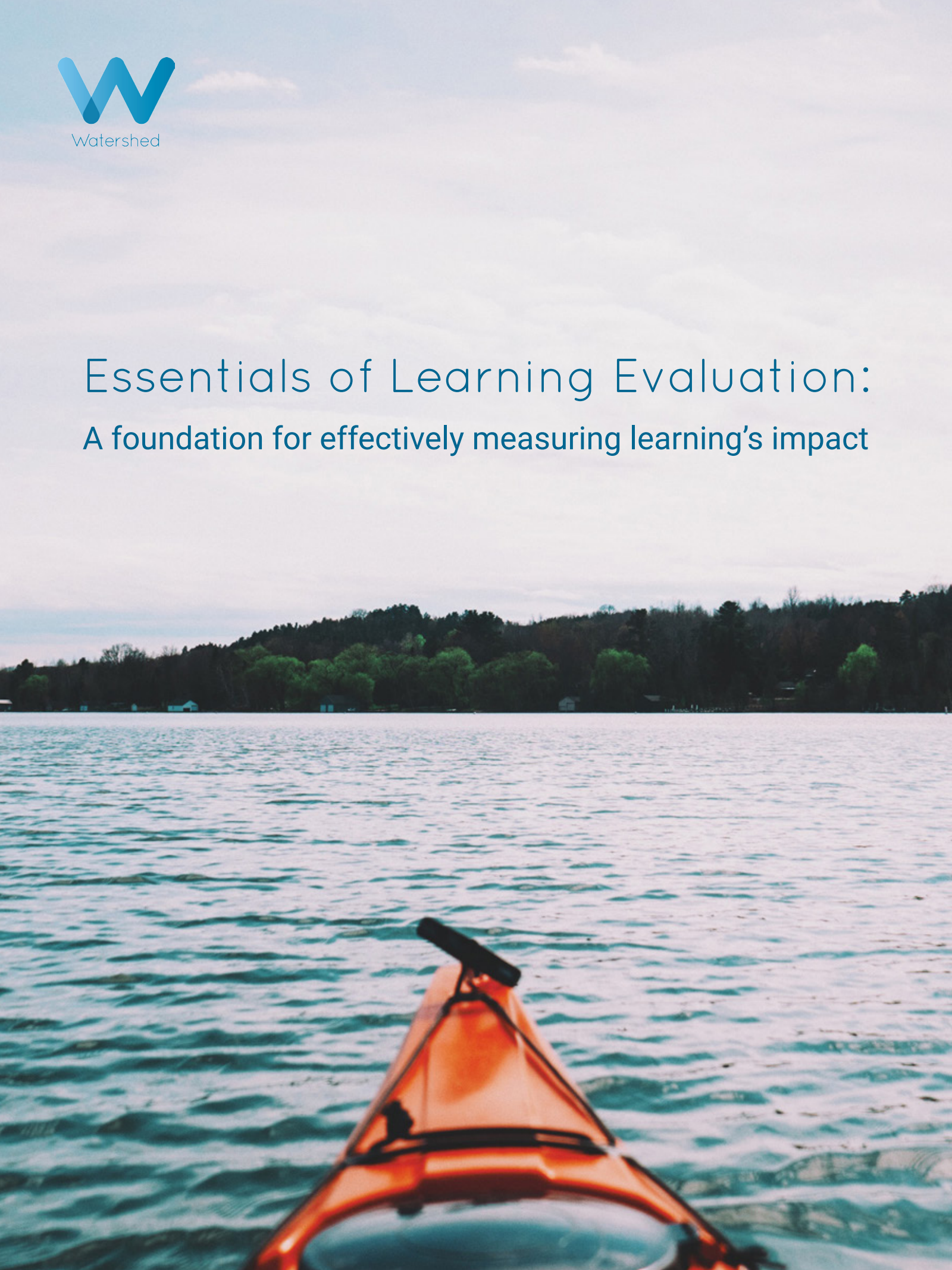
Watershed

Essentials of Learning Evaluation:

A foundation for effectively measuring learning's impact



Essentials of Learning Evaluation:

A foundation for effectively measuring learning's impact

Table of Contents

Introduction

Knowing how to effectively aggregate and evaluate all types of learning is a priority for many learning and development (L&D) professionals. According to a 2015 study from Bersin by Deloitte[1], L&D leaders are reporting more demand from their organizations to use data to demonstrate the impacts of learning programs.

But due to certain challenges—such as employees using multiple avenues for learning—many L&D professionals are finding it difficult to effectively measure and communicate the value of their learning programs.

Technologies are becoming available that enable L&D professionals to aggregate, evaluate, and measure many types of learning data in one place. That means it's learning's turn to prove its worth with evidence-backed facts and figures—which we believe will transform learning and, ultimately, transform businesses. We've got an exciting time ahead!

This whitepaper is ideal for:
L&D professionals who are new to learning evaluation and measurement
those who need a refresher on learning evaluation theories
those who need help deciding which learning evaluation theories best suit their needs
those interested in understanding how technology, xAPI, and measurement fit into learning evaluation

Key Takeaways:

This whitepaper:

- explores four established models of learning evaluation and their usefulness to L&D departments,
- reviews emerging technologies that can support learning evaluation as well as relevant technological developments in other sectors, and
- explains a seven-step guide by Watershed that will help to implement an evaluation process to measure the impact of learning programs.

Learning Evaluation Theory

Kirkpatrick

We'll start with the Kirkpatrick Model's four levels of evaluation as the most ubiquitous and well-known learning evaluation model. Since being defined in 1959, this model has had its fair share of both supporters and critics.

We'll look at some of the alternative models later in this whitepaper.

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 thatÕs 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 customersÕ 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 itÕs 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.

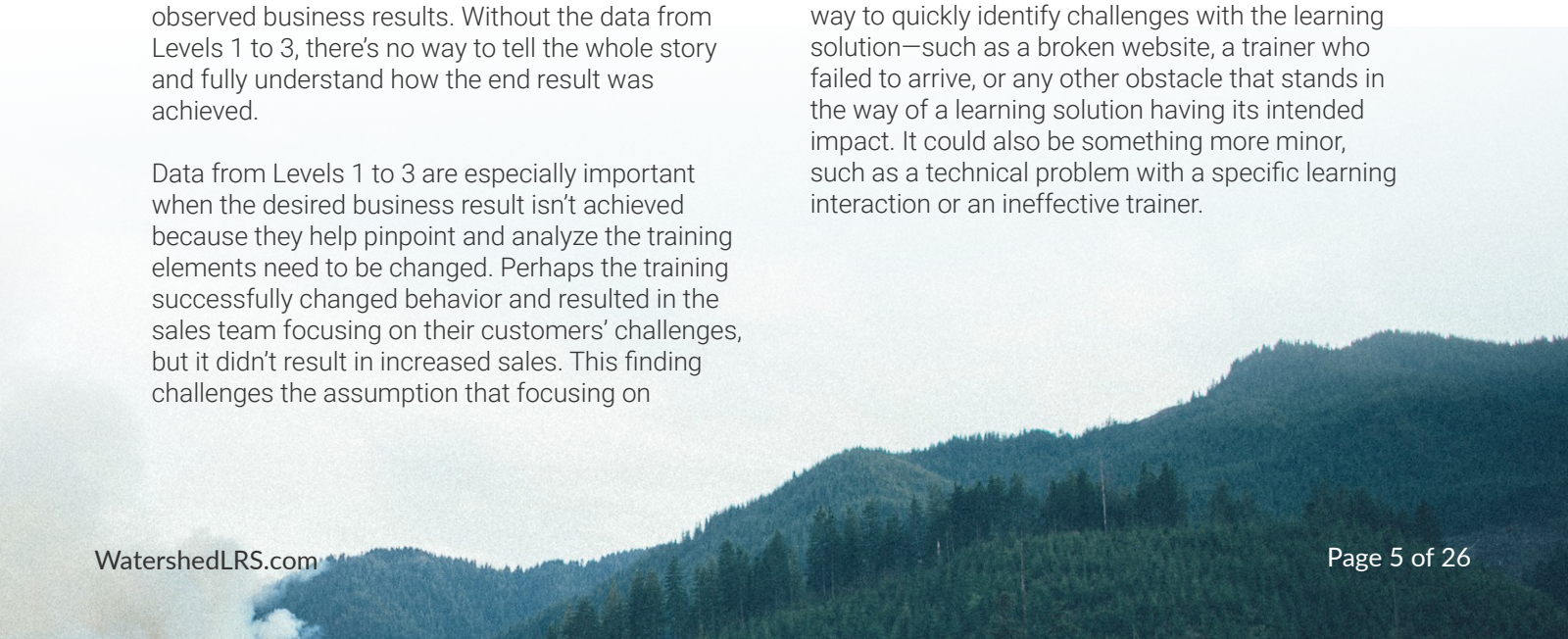
Investments in L&D are intended to drive positive business results.

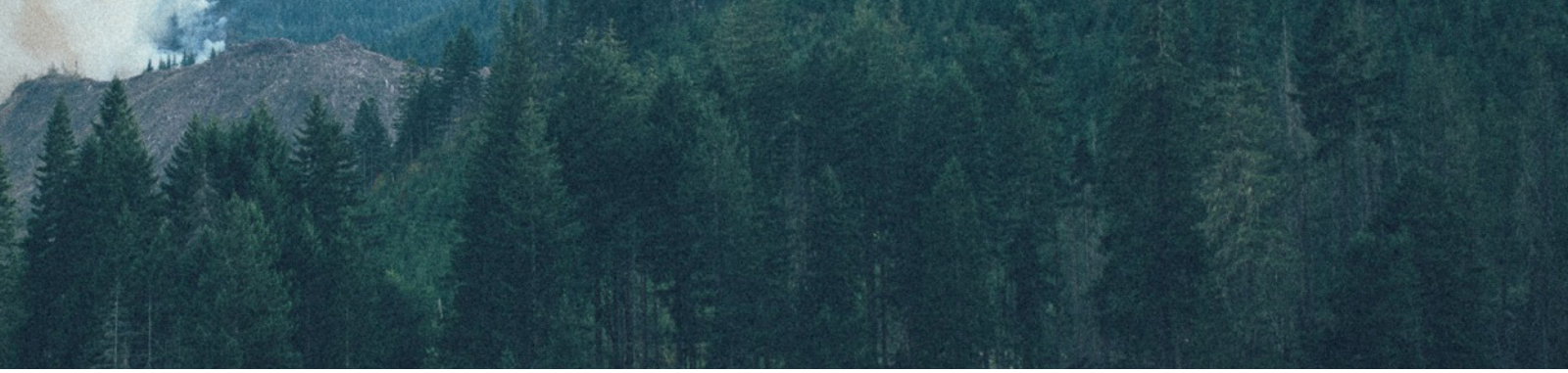
Is Level 1 Meaningless?

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[3]. 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 went wrong (or right).

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.

Kaufman's Five Levels of Evaluation

Kaufman's Five Levels of Evaluation is a reaction to and development of the Kirkpatrick Model's four levels. Where Kirkpatrick divides evaluation by type of impact, mainly to the learner, Kaufman's model evaluates the impact on different groups.

Kaufman's main developments from Kirkpatrick are:

- the splitting of Level 1 into input and process,
- the grouping of Levels 2 and 3 under the "micro" level, and
- the addition of fifth level, mega.

Kaufman also sees Kirkpatrick's model as being restricted to training delivery, while his own model considers both delivery and impact.

One interpretation of Kaufman's levels is summarized in Table 1, including the corresponding Kirkpatrick levels. (Note: This is not how Kaufman presents the final form of his five levels. We'll explain why later).

Kaufman	Kirkpatrick	Explanation
Input	1a	*Resource availability and quality* These are training materials, digital resources etc., used to support the learning experience.
Process	1b	*Process acceptability and efficiency* This is the actual delivery of the learning experience.
Micro	2 and 3	*Individual and small group payoffs* This is the result for the 'micro-level client' (normally the learner). Did the learner 'acquire' the learning? Did he or she apply it on the job?
Macro	4	*Organizational payoffs* This is the result for the 'macro-level client', the organization, and includes evaluation of performance improvement and cost benefit/cost consequence analysis.
Mega	n/a	*Societal contributions* This is the result for the 'mega-level client', either society as a whole or a company's clientele.

Input and process

The division of Kirkpatrick's Level 1 into input and process is perhaps the most practical and useful of Kaufman's suggestions. In a world that allows quick and easy access to websites—such as Google, Wikipedia, and YouTube—the availability and quality of web-based resources are becoming increasingly important evaluation factors. Different types of questions need to be asked when evaluating resource availability versus delivery, so it's helpful to think about them separately. Focusing on resource availability may be seen similarly to our suggested introduction of a level zero to Kirkpatrick, evaluating any informal learning that's happening socially or in the workplace. It's important to consider all available resources, not just those formally created within the organization.

Kaufman also replaces Kirkpatrick's measure of learner satisfaction with the learning experience, looking directly at learning resources and delivery themselves. It's helpful that Kaufman recognizes that, while input from learners is important when evaluating these elements, it's not the only source of data.

Micro-level evaluation

The grouping of Kirkpatrick's Levels 2 and 3 is less helpful, as learning and job performance can and should be evaluated separately. While we can't see inside the learner's brain, good assessments and simulations can capture data about learning. We can then track job performance to evaluate whether that learning has been correctly applied in the workplace.

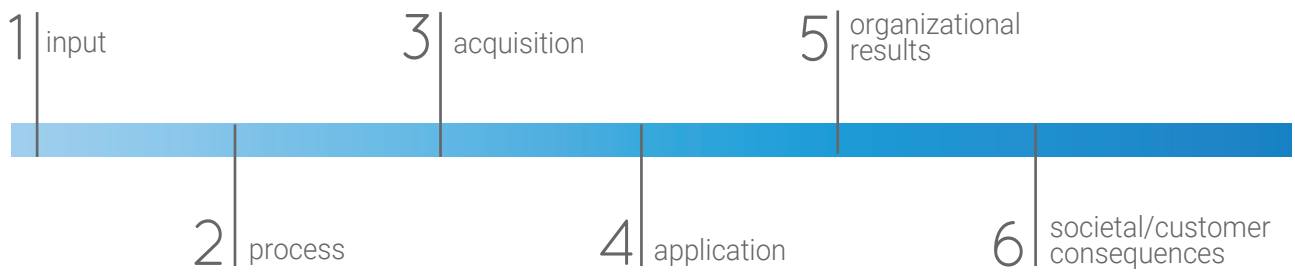
Having this evaluation data is important because it will determine the best way to resolve any issues. For example, the solutions to learners failing to apply their learning in the workplace are different from the solutions to learners failing to learn in the first place.

Six levels?

In Kaufman's final presentation of his five levels of evaluation, he attempts to mirror Kirkpatrick's levels, presumably to cater to those familiar with Kirkpatrick. This results in Kaufman keeping input and process together as Levels 1a and 1b of his model. At the same time, he keeps Kirkpatrick's Levels 2 and 3 separate, but titles them both "micro-level." This attempt at continuity with Kirkpatrick is understandable, but confusing.

Therefore, it may be more practical to think of Kaufman's model as having six levels and remove the mega/macro/micro terminology as illustrated in Figure 2.

Fig 2. Six Levels of Kaufman



Mega-level evaluation

Alongside the confusing terminology, the additional requirement to evaluate societal consequences and customer benefits make Kaufman's model less practical than Kirkpatrick's. We might be able to gather some anecdotal evidence about societal and customer impacts, but getting robust data at such a high level is often not feasible.

While it's helpful to consider the impact of learning on customers and society in some contexts, this evaluation often can be included in the business goal that the learning is expected to achieve. For example, if the learning is expected to improve sales, more customers will benefit because they're using your wonderful product. It's not necessarily helpful to evaluate that customer benefit separately from achievement of the business goal, though. Even when the goal is something such as "improving customer satisfaction," it doesn't need to be seen as a separate level from business results.

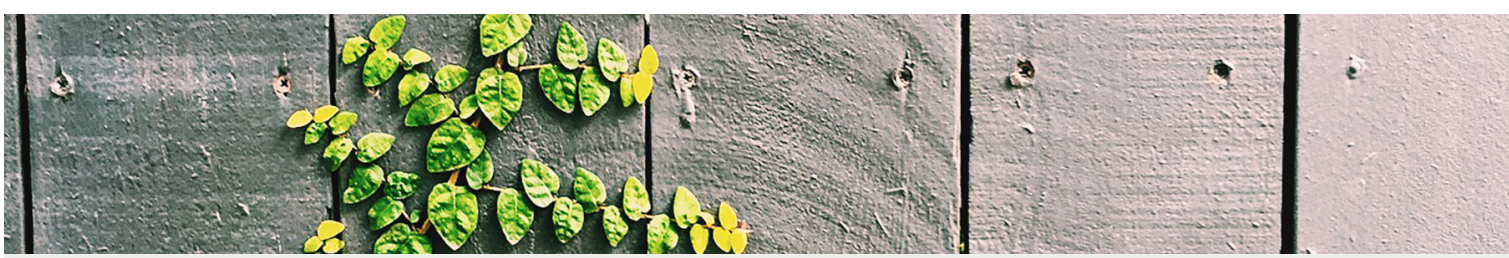
Just training

Kirkpatrick's original model was designed for formal training—not the wealth of learning experiences that happen in our organizations today. Kaufman's model is almost as restricted, aiming to be useful for "any organizational intervention" and ignoring the 90 percent of learning that's uninitiated by organizations. Further, it's hard to see how Kaufman's model is any better at tracking non-training interventions than Kirkpatrick's model.

In practice, Kirkpatrick is often applied in contexts outside of formal training. While the model was designed with formal training in mind, most L&D practitioners are pragmatic enough to reinterpret the model for their own particular contexts. We recommend this approach with any evaluation model; there will always be bits that work and bits that don't in any given context.

Kaufman vs. Kirkpatrick (our opinion)

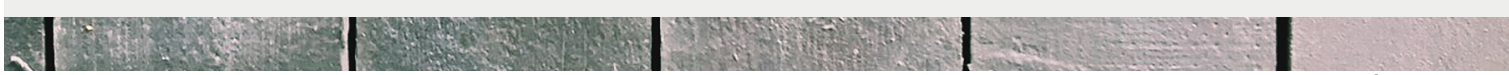
Kaufman's model provides some useful lessons that you can incorporate into your organization's learning evaluation strategy, but we don't recommend taking Kaufman's approach verbatim. In particular, the most helpful points are the division of resources from delivery and the move away from learner satisfaction. The least helpful facets are the addition of societal consequences and the overly complex terminology.



Kaufman Summarized

- Kaufman's model builds on Kirkpatrick's model and mirrors its four levels.
- Kaufman splits Kirkpatrick's Level 1 into "input" and "process."
- Kaufman adds a fifth level, which evaluates the results for society and the customer.
- Kaufman positions his model as more practical than Kirkpatrick's model (we argue this isn't accurate in practice.

While numerous evaluation models offer variations on Kirkpatrick and provide useful recommendations, we can't explore them all in this whitepaper. Instead, we'll look at two alternative models that take completely different approaches to Kirkpatrick.



Brinkerhoff's Success Case Method

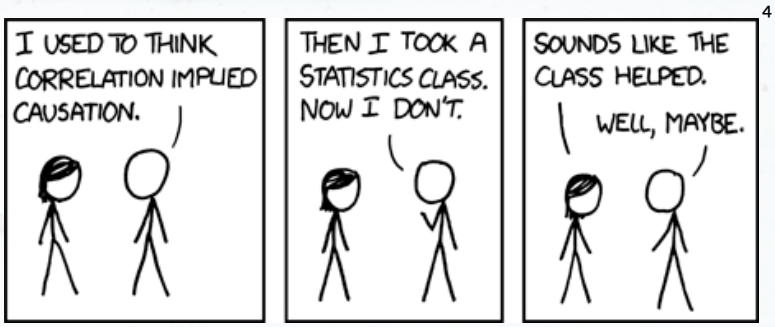
Kirkpatrick's evaluation model is normally implemented following quantitative research methods. Survey and assessment data are captured, aggregated numerically, and evaluated. By contrast, Brinkerhoff's Success Case Method focuses on qualitative analysis—crafting stories from discussions with a small number (i.e., about two to six) of affected parties.

Brinkerhoff's model isn't restricted to learning. It can be used to analyze any major business change, such as the purchase of new equipment or implementation of a new process. It's based on the assumption that any initiative, no matter how successful or unsuccessful, will always include some success and some failure. It seeks to uncover the most impactful successes and failures of an initiative and then tell the stories behind them, backed by evidence. Your organization can use these stories to learn how to be more successful in the future.

Part 1: Identify cases

The first stage of the Success Case Method identifies an initiative's success and failure. The beginning of this step is gathering the kind of quantitative data and surveys that might be collected as part of a Kirkpatrick-style model of evaluation. Data is collected to illustrate how successfully an initiative's business goal has been met across an organization. This data is used to find the outliers—cases where the initiative has been particularly successful or unsuccessful—in order to study them in more detail.

For this reason, the Success Case Method can be used alongside, rather than in place of, a Kirkpatrick-style evaluation model. We're advocates of both quantitative and qualitative research to support decision making. Ongoing quantitative analysis of big data provides up-to-the-minute indicators of your initiative's success as well as learning- and performance-related trends within your organization. Qualitative analysis enables you to dig into these results and explore their potential causes. You may have heard the saying "Correlation does not imply causation, but it does waggle its eyebrows suggestively and gesture furtively while mouthing 'look over there'." The kind of qualitative analysis suggested by the Success Case Method helps you to "look over there."



[4]

Part 2: Investigate further

The second stage of the Success Case Method involves interviewing people involved in these cases (i.e., the most and least successful stories). These interviews first seek to establish if there's sufficient evidence to verify each story. Once verified, the interviewer will gather further details and facts to produce a small number of comprehensive, evidence-based stories.

In most instances, there will be many more interviews than actual success cases. The stories told should cover all four of Kirkpatrick's levels of evaluation with evidence at each level to show the progression from the learning experience to learning and job application to business impact. These stories are then shared with the organization to applaud success and apply lessons learned to help improve and ensure the success of future initiatives.

Limitations

The Success Case Method focuses on the most and least successful cases resulting from a particular initiative. It's not designed to help judge the overall success of an initiative, but rather to learn from the most and least successful cases. We're not sure if Brinkerhoff would agree, but we recommend using his method alongside other evaluation methods to paint a full picture. Furthermore, the Success Case Method is:

- a manual activity that must be repeated every time evaluation data needs to be updated,
- effective within pilot programs to learn lessons from experimentation, and
- less useful for ongoing analytics where automated data collection and analysis becomes significantly faster and cheaper.

That said, the Success Case Method, or other qualitative methods, can be used to dig into unexpected changes in ongoing analytics data.

Brinkerhoff Summarized

This method focuses on qualitative data.

The goal is to create evidence-backed stories about the most and least successful cases from an initiative.

The method shouldn't be used on its own. Rather, use it to dig deeper into the results of quantitative analysis.

Because the Success Case Method is designed for the one-time study of an initiative's most and least successful cases, it's not useful in identifying overall success or for ongoing analytics.

Anderson's Value of Learning Model

A more recent and lesser-known model, Anderson's Value of Learning was published by the Chartered Institute of Personnel and Development and is based on research from the United Kingdom's University of Portsmouth in 2006. It aims to address two challenges:

Evaluation Challenge

Many organizations report that they struggle to do evaluation well.

Value Challenge

Organization leaders often require evidence showing the value of learning and training as well as cost-effective deployment of resources.

The Value of Learning Model is a three-stage cycle that's intended to be applied at the organization level, rather than for specific learning interventions.

Stage 1: Determine current alignment against strategic priorities.

This stage evaluates how closely learning in your organization is aligned to your organization's strategic priorities—such as driving sales, increasing production, or reaching a new market. To what extent is your learning strategy supporting that focus? Achieving high alignment requires knowing your organization's strategic priorities and developing a learning strategy that supports those priorities.

Stage 2: Use a range of methods to assess and evaluate the contribution of learning.

This stage evaluates the contribution of learning to your organization via a range of measures. The model doesn't detail exactly what these measures are, but outlines four areas of evaluation that should be covered.

- **Learning function measures** - How effective and efficient is the learning function within your organization? This includes the learning team and anybody else involved in supporting learning in the organization.
- **Return on expectation measures** - To what extent have the expectations of particular learning programs and interventions been met? For example, if you implemented a program to reduce a particular business process by two days, you'd use this measure to determine if you were successful in actually reducing this process by two days.
- **Return on investment measures** - How much did particular learning programs and interventions cost, and how does that compare to revenue generated and/or costs saved?
- **Benchmark and capacity measures** - How do learning processes and performance compare to internal or external standards? This evaluation should be monitored continually in order to show progress or decline compared to earlier data.

Stage 3: Establish the most relevant approaches for your organization.

The model recognizes that organizations differ and, as a result, that the measures outlined in the second stage will be more or less important depending on the specific needs of an organization. For instance, some organizations already recognize that learning is important and want to ensure that they're investing in the right initiatives, but other organizations need to be convinced. Similarly, certain L&D teams focus on the short-term benefits of learning, whereas others are more focused on the long-term benefits.

That's why it's important to choose a blend of metrics that best suit your organization's needs. Use Table 2 to determine which measures mostly likely to align with your organization's needs.

Table 2. Anderson's Value of Learning Model

	Senior management trust in the learning contribution	The organization requires learning value metrics
Emphasis on the short-term benefits	**Learning Function Measures**	**Return on Investment Measures**
Emphasis on the long-term benefits	**Return on Expectation Measures**	**Benchmark & Capability Measures**

Focus on learning strategy

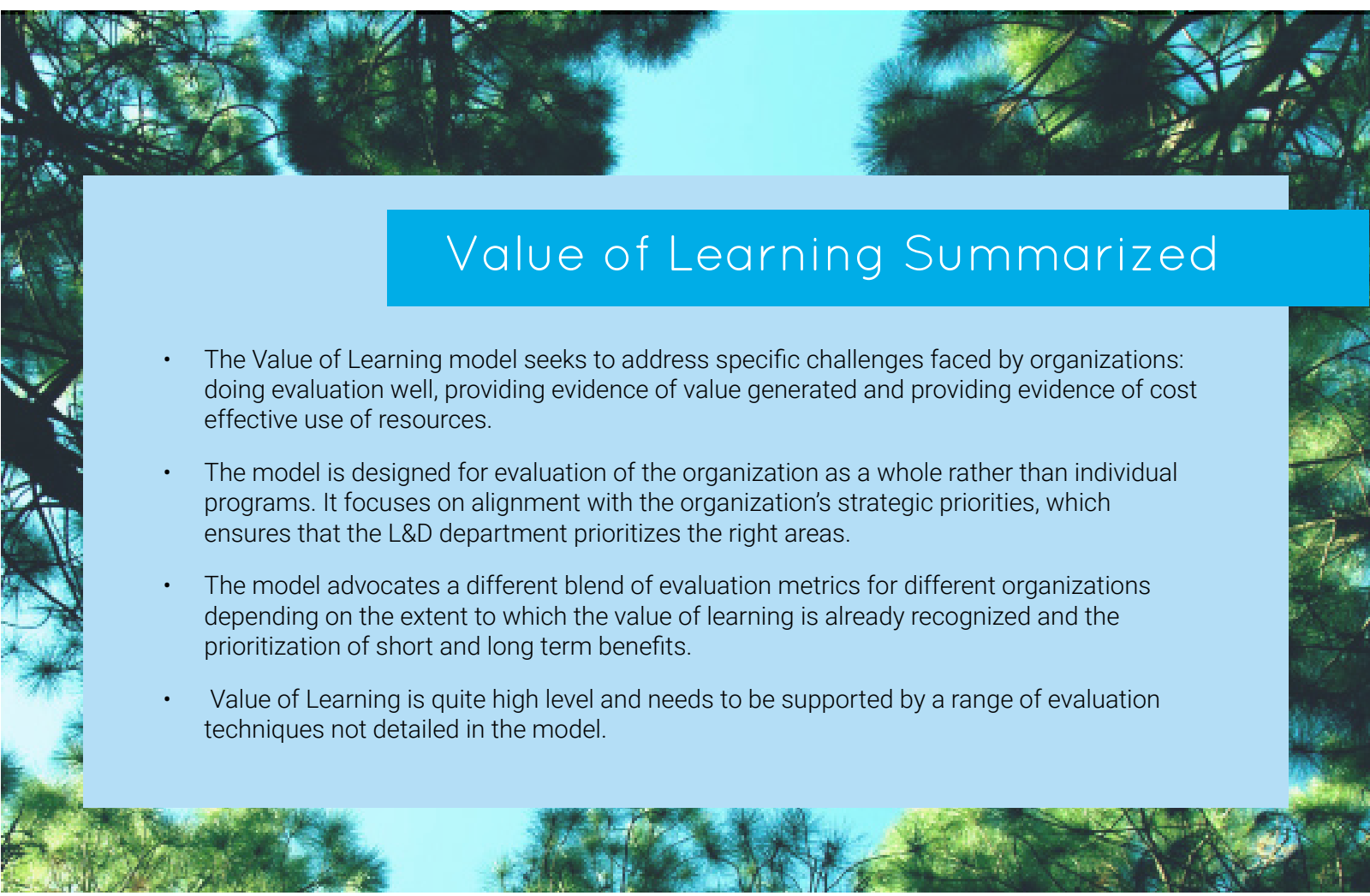
The Value of Learning model's focus on both learning strategy and alignment with an organization's strategic priorities is extremely helpful. For example, an organization might implement Kirkpatrick's model to evaluate a learning program designed to increase productivity of factory workers. The evaluation shows the program as a success if productivity increases. But what if that same organization already had surplus stock due to poor sales? That means the L&D department should have focused on upskilling the sales force to drive more sales rather than on increasing factory workers' productivity.

Assuming that same organization had sensible strategic priorities, an evaluation of how its learning program aligned with strategic priorities before starting a program would have identified that the program didn't match and resources would be better spent elsewhere. In other words, the Value of Learning model helps ensure you're not just delivering value, but you're also delivering that value where it's most critical to the organization.

Limitations

The Value of Learning model is deliberately high level and flexible, making it less useful in offering practical direction to specific evaluations. For example, the model encourages organizations to evaluate the effectiveness and efficiency of learning in the organizations as a whole, but it provides little direction as to how to measure either effectiveness or efficiency.

In practice, this model must be combined with other models, such as Kirkpatrick's model, to achieve the detail of individual learning initiatives in order to paint an overall picture of the effectiveness of learning in the organization.



Value of Learning Summarized

- The Value of Learning model seeks to address specific challenges faced by organizations: doing evaluation well, providing evidence of value generated and providing evidence of cost effective use of resources.
- The model is designed for evaluation of the organization as a whole rather than individual programs. It focuses on alignment with the organization's strategic priorities, which ensures that the L&D department prioritizes the right areas.
- The model advocates a different blend of evaluation metrics for different organizations depending on the extent to which the value of learning is already recognized and the prioritization of short and long term benefits.
- Value of Learning is quite high level and needs to be supported by a range of evaluation techniques not detailed in the model.

Learning Evaluation & Technology

The increasing demand for better evaluation and analytics in learning has largely been driven by better data in other business areas. For example, the marketing function used to be a black box where investments had unknown impacts. Now, marketers are able to easily see and measure the direct benefits and value of specific marketing investments.

Understanding the audience

Google Analytics is an example of a commonly used tool that provides detailed insights into website traffic and behaviors. This includes data about website visitors themselves, but more important, how visitors got to the website and then navigated through it. Marketers can clearly and accurately see which marketing efforts led to specific sales and leads, enabling marketers to put a dollar value on almost every marketing activity with a strong degree of confidence.

Understanding behavior

Technology also is being used to drive product direction. Netflix, for example, captures detailed data about their 65 million customers and their entertainment preferences to personalize viewing experiences. This includes creating 15 trailers for the same show with each trailer targeting different groups of Netflix users. These analytics also provide an accurate indication of whether or not a show will be watched by customers. This not only helps Netflix determine the best movies and shows to license for its customers, but also ensures Netflix isn't investing in the wrong types of entertainment.

So, what about learning?

How can you quantify the value of your learning program in relation to your business? How can you identify ahead of time which programs are likely to have an impact and which will fail? What technology exists to get and analyze that data?

For both examples discussed above, data must first be collected from a single source—whether it's a website connected to Google Analytics or the Netflix platform. In learning, we don't have one single source of data about the learning that's happening in an organization. Even if the organization has a single, well-used LMS, that LMS is probably only capturing data for a limited amount of the workforce's learning experiences. Most learning is happening elsewhere—such as intranets, simulators, social platforms, external websites, offline conversations, and job tasks—and capturing all the data from these disparate sources is a challenge.

We've got options

Fortunately, it's a challenge the learning technologies industry has been aware of and working on for quite some time. After several years of research and development, a community group focused on the learning technologies industry (including us) released an interoperability specification known as the Tin Can API (xAPI) in April 2013. This specification defines a common way for learning technologies to communicate data about learning experiences to a central Learning Record Store (LRS) for analysis. As more data sources become "xAPI enabled," it becomes easier for us to perform the evaluation methods we covered earlier in this whitepaper.

As mentioned in the section discussing Kirkpatrick, xAPI also helps evaluate even formal learning experiences in terms of learner behavior and business impact. Companies are using xAPI to track job performance alongside learning to explore the impact of their learning interventions.

Comprehensive analytics

The biggest change these technologies bring is the ability to quickly establish and easily track ongoing analytics. Without this technology, learning professionals would spend countless hours continually gathering, compiling, and updating data. Using an LRS and xAPI, data is collected and presented on an ongoing basis and with a Learning Analytics Platform, such as Watershed, learning professionals and senior managers can access up-to-date analytics.

Learning Evaluation in Practice

We've explored four learning evaluation models and several technological advances that are enabling better analytics. But what does all of that mean for your organization in practice? This section explores our end-to-end approach to practical evaluation with case study examples.

Watershed's Seven Steps of Evaluation

We recommend a seven-step process for evaluating L&D programs. Our approach to evaluation is a combination of Kirkpatrick's model and the Value of Learning model with lessons from Kaufman's model, the Success Case Method, and what we've found works in practice.

Fig. 4

Watershed's Seven Steps of Evaluation

1 Align

Identify project goals and evaluate alignment with strategic priorities (Value of Learning).

2 Define

Identify success metrics most appropriate to the organization (Value of Learning). Ensure that these metrics cover the complete story from learning to achievement of the project goals (Kirkpatrick and Kaufman).

3 Discover

(Optional) Identify what learning is already occurring in the organization that supports the projectõs goals. Research those activities and identify what works well and what doesnÕt (Success Case Method).

4 Design

Design the project itself, including determining how data relating to evaluation metrics will be captured, aggregated, and displayed/used. Consider use of dashboards for ongoing monitoring and reports for analysis at specific points in the project.

5 Monitor

Continually monitor success and progress toward the project goal and keep stakeholders updated. Make changes to the project as required in response to ongoing data.

6 Analyze

At the end of the project and/or at specific dates after implementation, analyze data in detail. Celebrate and share evidence of successes; document and share lessons learned.

7 Explore

(Optional) Research further into particularly successful and unsuccessful elements of the project to uncover more lessons learned (Success Case Method).

1 Align

Identify program goals and evaluate alignment with strategic priorities.

The most important step in learning evaluation is understanding the program's purpose and goals. This involves digging into why the program was created in the first place, identifying and talking to project stakeholders, and reflecting on what changes the program is supposed to bring about. In some cases, the aims of the program are obvious. In other cases, the underlying goals can take more work to uncover (perhaps the L&D team has simply been asked to "make a course on X").

Once the program's goals are defined, evaluate how these goals fit with the strategic priorities of the organization. This is a vital part of program evaluation. As we explored in the Value of Learning model white paper, a successful program that doesn't align with overall strategic priorities is a bad use of the organization's resources. Given that most L&D departments have limited time and resources, work that has the biggest strategic impact should be prioritized over those that make the most noise.

By the end of this step, you should have:

- identified your program's goals,
- evaluated how closely they align to strategic priorities, and
- determined whether or not the program will proceed.

2 Define

Identify success metrics most appropriate to the organization.

As we discussed in the previous section, the first step of evaluating your learning programs is aligning the program goals with strategic priorities. With the program goals aligned, they can be broken into individual evaluation metrics that will show the success of the program and monitor progress toward the program goal. To some extent, this step happens in conjunction with the discovery and design steps as the information gathered and decisions made in these steps affect one another.

When choosing success metrics, we recommend a collection of metrics that cover as many of the following five areas below.

1. Usage and quality of content
2. Delivery and discoverability of learning experiences
3. Improvements in knowledge and competency
4. Improvements in job performance
5. Overall achievement of the program goals

Each program goal should have one KPI that can be monitored throughout the program. These will be the most important metrics of your evaluation.

This list provides examples of evaluation metrics at each level:

1	Usage and quality of content	• Number of times content is accessed* • How much and which parts of video content are skipped • Attendance at events • Time spent • Difficulty and quality of particular assessment questions • Survey responses
2	Delivery and discoverability of learning experiences	• Number of times the experience is experienced* • Clicks or time taken to access the learning experience • Survey responses
3	Improvements in knowledge and competency	• Assessment/simulation scores showing progress before and after the learning experience.
4	Improvements in job performance	• Peer/manager/mentor observations • Customer feedback • Data from business systems used to do the work • Survey responses
5	Overall achievement of the program goals	• Number or value of sales made per month • Average time taken to produce each widget • Average customer satisfaction rating • Employee retention %

*Number of hits applies to both levels as the reasons for high or low numbers could relate to either quality or discoverability. This metric needs to be used in conjunction with other data or research.

The metrics used and how they are expressed will depend on both the program and the organization. This is particularly significant for the metrics relating to the overall program goal. For example, in an organization interested in short-term benefits that has confidence in the value of learning, the metrics might stand on their own. In an organization that's more interested in long-term benefits and with low confidence in the value of learning, the metrics should be expressed in relation to industry benchmarks (see the Value of Learning model).

Whatever metrics you choose, be sure to measure them over a period of time. It's important to evaluate whether or not learners are remembering their learning and persisting in their changed behaviors.

For each metric, determine threshold values that represent expected, below expectations, and above expectations. This is important to determine your level of success. These thresholds may change, for example, as your organization improves. Remember to link these thresholds to applicable industry benchmarks or levels of return on investment/expectation.

In addition to creating metrics, you should define a set of quality criteria for your materials and resources. These can be applied before any learners interact with them and should reflect both production quality (e.g., are images used appropriately?) and pedagogy (e.g., does the solution include reminders to aid remembering?).

Once you're satisfied with your chosen metrics (and how they're expressed), review them with relevant stakeholders to ensure the metrics provide all the information that's needed.

By the end of this step, you should have defined a set of metrics to monitor and analyze the program. Expect these metrics to evolve as you discover existing learning within your organization and then design the processes to capture, analyze, and present these metrics.

Example 1: Devereux

Devereux planned to improve treatment outcomes by observing employees in their daily work. They wanted to test the theory that observations would improve performance. Their two key metrics were number of observations and patient risk scores.

Example 2: Pandora

Pandora's program aimed to increase product sales through the use of formal and social learning experiences via their LMS. Their two key metrics were product sales and interactions with the LMS.

3 Discover *(optional)*

Identify what learning is already occurring in the organization that supports the program's goals. Research those activities and identify what works well and what doesn't.

As we've already discussed, the first two steps for effective learning evaluation are aligning program goals with the business and defining success metrics. The third step for learning evaluation is discovering what learning programs are already working well in the organization.

Before designing your program, it's important to research activities and/or learning that may already be happening in your organization, and to identify which of these activities are supporting or preventing the project's goals. Keep in mind, this type of investigation tends be relatively manual, so it might be appropriate to apply the Success Case Method.

While it's helpful to ask learners to record their informal learning experiences, many people aren't usually aware of just how much they've learned or grown as a result of everyday interactions. Peer or manager assessments and performance metrics can uncover changes in competency that would otherwise have been missed.

If you discover (or are already aware of) existing learning activities that might contribute to program goals, consider including evaluation/monitoring metrics for those activities in addition to the program's formally created/curated activities.

By the end of this step, you should have:

- identified formal and informal learning within your organization that relates to the program's goals,
- evaluated the effectiveness of these learning experiences, and
- determined the extent to which the learning is positive (are people learning the right thing?).

Design

Design the program itself—including determining how data relating to evaluation metrics will be captured, aggregated, displayed, and used.

4

Design the program with a set of evaluation metrics you want recorded throughout the program. The program design should include a plan for how these metrics will be captured, aggregated, and then displayed to relevant people.

During Step 4, assess the feasibility of capturing each metric. You'll most likely need to adjust some methods to make capturing metrics easier for your particular goals. Because of financial and labor constraints, some metrics might be deprioritized or dropped entirely. Use the wish list of metrics defined during Step 3 to refine which ones are most useful. You can identify which metrics are the best options by comparing the cost of capturing, aggregating, and presenting each one.

As part of your program design, consider if you should include A/B testing—comparing two versions of a particular element to see which one performs better. You can then use your evaluation metrics to judge which element(s) was most effective and follow that approach when you go live with the full program.

Now that you've designed how you'll capture, aggregate, and display metrics, you can launch your program. We recommend creating dashboards to monitor metrics throughout the program.

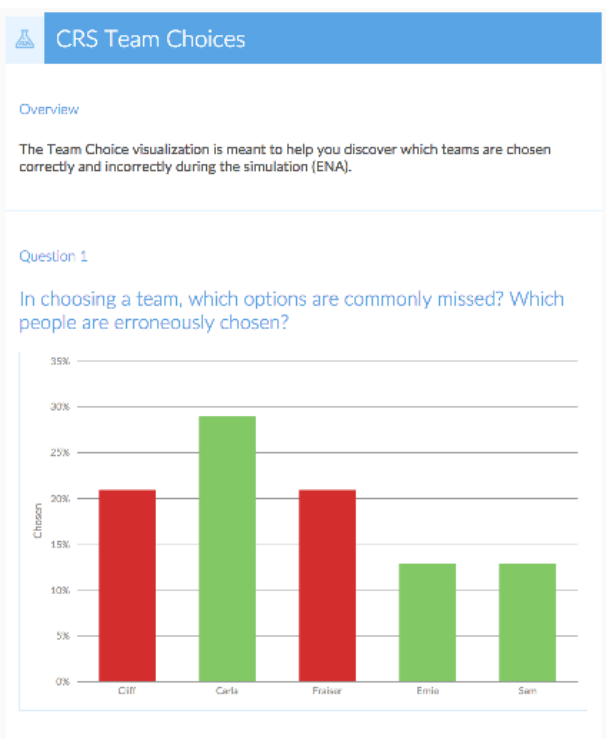
By the end of this step, you should have:

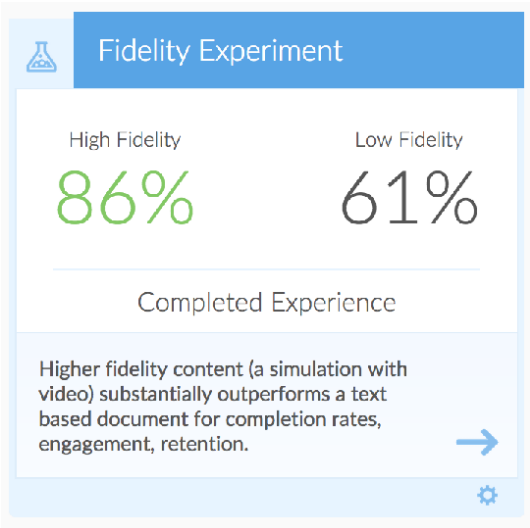
- identified any problems early on, and
- resolved those problems.

Example 3: Catholic Relief Services (CRS)

CRS used Watershed to evaluate the effectiveness of their training program on teams working in disaster areas. They pulled data from existing training systems, such as their LMS, and collected data about performance in the field via a mobile app that provided a performance observation checklist.

This data was displayed in a range of dashboards and reports, such as the one to the right.





Example 4: AT&T

AT&T tested a new approach to compliance and ethics training. This involved two levels of situational simulations that were randomly assigned to learners who chose to participate.

Data was aggregated from the simulation, assessment, and training path systems into an LRS. The data from both the test group and the control group were compared to assess learner engagement and retention.

5 Monitor

Continually monitor success and progress toward the program goal and keep stakeholders updated. Make changes to the program as required in response to ongoing data.

As discussed in Step 4, your design should include dashboards to help you monitor metrics related to your program. Check these dashboards regularly to make sure no issues arise within your program. And if an issue does arise, address it quickly.

Let's say you're rolling out training to 5,000 learners to improve their job performances in a particular area. You've included pre- and post-assessments of the training and are using a dashboard to monitor the change in scores before and after training. After a few days, 750 learners have completed the training as a pilot, but the dashboard isn't showing statistically significant change in assessment results. This data allows you to quickly postpone the main block of training while you investigate what's not working (or why the assessment is bad). You can then improve and rerun the pilot program before you waste the time of 4,250 other learners. Don't wait until the end of the program to evaluate and make improvements.

This step is ongoing throughout the program. You should identify any problems that arise early on and implement the required changes to fix those problems.

Example 5: CUES

CUES is using Watershed to evaluate the use of learning resources on their website. Their dashboard allows them to quickly identify any changes in use of different resources.

Analyze

At the end of the program and/or at specific dates after implementation, analyze data in detail. Celebrate and share evidence of successes; document and share lessons learned.

6

The dashboards set up during Step 4 will help you monitor the success and progress toward the metrics most appropriate to the organizational goals.

Dashboards are great for quick, accessible information, but at the end of the program (or at specific review points) you'll want to dive deeper into the data for a more thorough analysis. That's why you'll need reports to provide more in-depth data. Your reports should enable you to tell the whole story of what happened, supported by evidence, and pinpoint why your program was successful or unsuccessful.

Use this data to show the program's success and why it's a worthwhile investment. For instance, you might include industry benchmarks, return on investment, return on expectation information, or just the story of the data itself. Be sure to document and emphasize the successes to reinforce the importance of L&D in your organization while making the business case for investment in future programs. If areas of the program haven't shown success, explore data to pinpoint what needs improvement for future programs.

By the end of this step, you should have:

- identified the extent of the program's overall success,
- determined the reasons for the program's success, and
- compiled positive success stories and lessons learned for future programs.

7

Explore *(optional)*

Research further into particularly successful and unsuccessful elements of the program to uncover more lessons learned.

Dig deeper into success stories or lessons learned by using more qualitative research methods, such as the Success Case Method. You may not want to do this for every program, but it's worthwhile for larger programs and for your first few programs implementing this model. As set out in the Success Case Method, identify the program's most significant successes and failures, research the reasons for these outcomes, and then document the stories.

By the end of this optional step, you should have:

- documented several of the program's most and least successful stories,
- identified even better stories that promote the program's success, and
- gained a better understanding of how to improve future programs.

Summary

Evaluation is a vital part of learning programs, but it's often overlooked. Evaluation must be made at all stages of the journey—from learning resources to delivery and application in the workplace to the business goal, which must be aligned with your business's strategic objectives. Remember, your evaluation metrics will vary depending on your organization's attitude to learning. Your evaluation data should provide insights to improve your learning strategy, and you can get even deeper insights by further research into the most significant successes and failures of your programs.

Technology is transforming every aspect of our organizations, and learning is no exception. The growth of xAPI (i.e., Tin Can) is enabling organizations to easily collect and analyze learning data from many disparate sources.

Our seven-step model provides practical guidelines that you can apply in your organization for all your learning programs. Following these steps will help you identify success and areas for improvement throughout the entire duration of your program. Speak to us about implementing these steps in your organization for your next program.

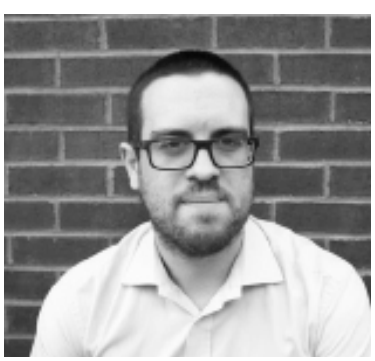
Special Thanks

Special thanks to Will Thalheimer, PhD, learning and performance consultant, who reviewed this whitepaper prior to publication. Dr. Thalheimer made a number of suggestions and comments that have been incorporated.

That doesn't mean Dr. Thalheimer necessarily endorses any or all of the arguments put forward in this whitepaper. In fact, his main criticism of this whitepaper is that we over-focus on business goals and forget that the primary focus of learning professions is learning rather than performance. Our evaluations shouldn't be so focused on business goals that we forget to evaluate learning. He has defined seven Training Maximizers that are meant to help ensure our own learning solutions will achieve maximum results.

At Watershed, our take is that both learning and business goals are important. Learning activities should work toward a business goal, but that doesn't make evaluating the learning itself unimportant. We need to evaluate at every level to tell the whole story and identify where our learning programs can improve.

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 LRS, 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.

Watershed

Sources

[1]Rochelle, Michael. "Learning and Development: Framework for Excellence." www.brandonhall.com. Brandon Hall Group, May 2015. Web.

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

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

[4]Munroe, Randall. "Xkcd: Correlation." Xkcd: Correlation. N.p., n.d. Web. 26 Aug. 2015. https://xkcd.com/552/.

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

Kaufman, R., Keller, J. and Watkins, R. (1996), What works and what doesn't: Evaluation beyond Kirkpatrick. Nonprofit Management Leadership, 35: 8–12. doi: 10.1002/pfi.4170350204

Brinkerhoff, Robert O. "What Is the Success Case Method and How Does It Work?" The Success Case Method: Find out Quickly What's Working and What's Not. San Francisco, CA: Berrett-Koehler, 2003. 1-26. Web. www.bkconnection.com/static/The_Success_Case_Method_EXCERPT.pdf. 24 Aug. 2015

Gram, Tom. "Evaluating with the Success Case Method." Performance X Design. N.p., 24 Feb. 2011. Web. 24 Aug. 2015.

Anderson, Valerie. "The Value of Learning: A New Model of Value and Evaluation." Royal Charter (2007): n. pag. Chartered Institute of Personnel and Development. Web. www.cipd.co.uk/NR/rdonlyres/94842E50-F7754154-975F-8D4BE72846C7/0/valoflearnnwmodvalca.pdf. 26 Aug. 2015

www.watershedLRS.com