The Learning-Transfer Evaluation Model:

Sending Messages and Nudging Evidence-Informed Thinking to Enable Learning Effectiveness

Aligned with LTEM Version 13

by Will Thalheimer, PhD, MBA

The Learning-Transfer Evaluation Model:

Sending Messages and Nudging Evidence-Informed Thinking to Enable Learning Effectiveness.

By Will Thalheimer, PhD

Citation:

Thalheimer, W. (2024). *The Learning-Transfer Evaluation Model: Sending messages and nudging evidence-informed thinking to enable learning effectiveness.* Available at https://WorkLearning.com/LTEM.

LTEM Version 13 incorporates many significant changes, including wording improvements at all eight tiers. Also, LTEM Tier-4 Knowledge has been divided into two levels: 4A, representing "Foundational Concepts," those helpful in making work-related decisions, and 4B, representing "Trivial Knowledge," information not central to work decision. Finally, the framework has been expanded to two pages, with the second page listing, "Helpful Hints on How to Use LTEM."

I welcome comments and suggestions at info@worklearning.com!

Special Thanks:

To Julie Dirksen, Clark Quinn, Roy Pollock, Adam Neaman, Yvon Dalat, Emma Weber, Scott Weersing, Mark Jenkins, Ingrid Guerra-Lopez, Rob Brinkerhoff, Trudy Mandeville, Mike Rustici, Mike Saunderson, Danielle Dziurun, Shilpa Subramaniam, Ruhi Grewal, Elham Arabi and many other people—including those who participated in my workshop, *The LTEM Boot Camp*. Without the thoughtful and extensive feedback I have received over many years, neither the model nor this report would be any good.

The Learning-Transfer Evaluation Model:
Sending Messages and Nudging Evidence-Informed Thinking to Enable Learning Effectiveness.

by Will Thalheimer, PhD

LTEM Overview

LTEM is a learning-evaluation framework designed to help learning professionals consider a broad range of options for measuring learning outcomes. It is specifically designed to encourage evidence-informed thinking, to infuse learning wisdom into learning evaluation, and to support us in creating learning programs that transfer to work performance.

LTEM is also designed to discourage poor learning evaluation practices, or practices that are insufficient to validate learning success. For example, LTEM signals that measuring attendance, completion rates, and learner activity are all insufficient in validating learning—because, even when these results look good, it can still be the case that learners didn't learn, or they learned bad information, or they became demotivated to apply what they learned.

LTEM was in design throughout 2016 and 2017 and went through 11 iterations before being published in 2018. Many people provided input, criticism, and advice during the development process. Version 11 was published in February of 2018 and Version 12 was published in May of 2018.

From 2018 and through the next six years, LTEM continued to rise in popularity throughout the world, often recommended by the most enlightened learning-and-development thought leaders.

LTEM was developed in response to the well-known limitations of the Kirkpatrick-Katzell Four Level Model of Evaluation. The four-level framework was devised by Raymond Katzell and refined with labels for each of the four levels by Donald Kirkpatrick in his often-cited articles of 1959 and 1960.

LTEM has significant advantages over the four-level framework and other learning-evaluation frameworks. Foremost, it brings learning wisdom into our

learning-evaluation deliberations. Instead of having just one container for learning—what the Four-Level Model refers to as "Level 2 Learning"—LTEM separates learning measurement options into Knowledge, Decision Making, and Task Performance. This revolutionary framing enables learning professionals to distinguish between (and target) more work-relevant measures and avoid less relevant measures like knowledge checks.

LTEM is designed with numerous critical features that enable it to better support learning teams in engaging in learning evaluation and using LTEM as an integral part of their learning-development process.

Like all models, LTEM has limitations. Specifically, it is designed with a focus on learning transfer and is not appropriate for everything we learning-and-performance professionals might want to measure. Still, LTEM is currently the most comprehensive learning-evaluation framework and can be augmented with the LEADS framework—Learning Evaluation As Decision Support—which I've introduced formally in the book *The CEO's Guide to Training, eLearning & Work: Empowering Learning for a Competitive Advantage*.

To enable clear communication, LTEM should be written as "LTEM" and pronounced as "L-tem."

Version 13. Published October 2024.

LTEM Version 13 adds some refinements, specifically these improvements:

- The eight tiers remain the same, but Tier-4 Knowledge is now divided into two levels: 4A (Foundational Knowledge) and 4B (Trivial Knowledge). This distinction is critical, as will be explained later.
- The wording in each tier has been refined from earlier versions based on how LTEM has been interpreted over the past six years.
- LTEM is now designed to better meet accessibility standards.
- LTEM has a new, cleaner design.
- LTEM has been expanded, with a second page added to provide specific hints about how to use LTEM.

Uses for LTEM

LTEM is a conceptual model, designed to help learning professionals think effectively about learning measurement and learning practice. LTEM does not handcuff us by specifying a static set of metrics but enables flexibility by providing a robust framework inspired by the research on learning.

The following list provides 16 ways LTEM can be used.

1. Evaluation Gap Analysis

On our learning-and-development teams, we can use LTEM to map out the gaps between our current evaluation practices and our desired goals for evaluation. One learning team may find that they are primarily measuring completion rates (LTEM Tier-1) and using traditional smile sheets (LTEM Tier-3B) when what they really want to do is use performance-focused learner surveys (Tier-3A), scenario decisions (Tier-5), and measure transfer to work performance (Tier-7) for their two most important learning programs. Another learning team may find that they are measuring most of their programs at Tier-5 and Tier-6 and conclude that they are exactly where they want to be.

By finding the gaps between our goals for evaluation and our current practices, we can make plans to move toward our aspirations.

2. Brainstorming Metrics to Use

When we are evaluating a particular program, we can use LTEM to brainstorm possible metrics at each tier. Through this divergent thinking process, our team is likely to generate metrics we hadn't considered—enabling innovations that provide real value.

We must remember good brainstorming rules. Have individuals brainstorm on their own rather than having people brainstorm together. This increases the quality and breadth of the ideas generated while preventing groupthink.

3. Prioritizing Metrics to Use

Evaluating learning is an investment. It requires time, resources, and money. We can't use every evaluation metric we brainstorm. LTEM provides a nice conceptual map to guide us in thinking about how to prioritize.

4. Framing Evaluation Targets

Evaluation projects often get messy and require mid-course corrections. Where we're using more than one metric—say, we're using learner surveys, scenario questions, and capturing key performance indicators (KPIs)—LTEM can act as a visual guide for what we've agreed to accomplish. Moreover, the tier labeling can keep everyone clear about our evaluation targets. *"Okay, team, before we start our meeting today, let's remember that our plan is to measure at LTEM Tier-3 using performance-focused learner surveys, at Tier-5 using scenario questions, and at Tier-6 using a troubleshooting simulator. Are we still on board with this plan?"*

5. Nudging and Supporting Learning Design

LTEM can inspire more effective learning designs. For example, early in the learning-development process, we can brainstorm learning metrics and then prioritize them down to a set of "evaluation objectives." We might decide to use scenario questions presented to learners after a week's delay to determine if the learning program helped participants make work-realistic decisions based on what they learned—an LTEM Tier-5 metric. We might select a few other evaluation objectives as well, Tier 4A knowledge checks, and Tier-3A learner-survey questions.

After selecting our evaluation objectives, we can then craft our learning objectives, avoiding the faulty action-verb compilations that previously pushed us to low-level goals like having the learners "explain," "list," or "describe." Our team leader might say, *"Well, if we are going to be measured on the ability of the learners to make work-realistic decisions, we'd better give the learners more practice in making decisions."*

The beauty of this is that giving learners practice in work-realistic decisions is one of the best learning-design methods, certainly much better than low-level knowledge checks.

And, FYI, this is exactly what happened when Elham Arabi, in her doctoral dissertation research, introduced LTEM to a learning team focused on nurse professional development in a hospital. LTEM not only inspired better learning evaluation, but it also nudged the learning team to utilize more effective learning designs.

6. Credentialing Based on Course Quality

Some learning experiences are better than others. Some providers of learning programs produce more effective learning interventions. Should we treat one course like any other course? No! But we do. It's crazy, really, and it harms our work because our lack of awareness on course quality pushes us to poor decision making. Ideally, we would nudge our learners to take more rigorous courses rather than less rigorous ones. We would hire vendors that produce more effective courses. We would do more to ensure our own courses are more rigorously measured.

The problem is that validating courses and vendors takes too much effort. Indeed, given how hard it is to gather data on other people's programs, such an effort is next to impossible. But what if we could shortcut this process a little? LTEM enables a reasonable plan of action.

Let's take two courses. One is evaluated with traditional smile sheets (LTEM Tier-3B) and the other is evaluated by asking learners to perform the targeted work tasks (LTEM Tier-6) to demonstrate that they've learned. All other things being equal, it's likely that the course requiring more rigorous evaluation methods is better. First, it is asking more of the learners, which is almost certain to strengthen their learning. Second, by getting more rigorous feedback, the course owners are likely to have engaged in many rounds of improvement based on reasonably good learning metrics.

The bottom line is that we can use LTEM as a credentialing tool. It won't be as good as putting learners through a proper credentialing test—an unbiased test given by an independent sanctioning body. But, by examining the LTEM rigor of the learning metrics, we can get a reasonable idea about course quality.

7. Backward Chaining from Ultimate Goals

Learning is wicked complex. When we think about what is required, we'll notice a causal pathway from learning to results. So, for example, we need to first provide valid, relevant, credible content to learners. They must pay attention to it in ways that help them grasp the content. They must comprehend the key concepts and procedures fully and accurately. They must be motivated to apply what they've learned. They must remember the concepts and retrieve them in the right context when they need them in the workplace. They must make an honest attempt to apply the learning and overcome obstacles and perhaps

remain resilient over time. Only after all this might we find evidence that the learner is performing and producing results.

Look at all the things we might measure. Wow! And since this is a causal chain, if a key step fails, then the results will not meet our goals. If a learner is motivated but doesn't remember what they've learned, then they won't be able apply what they've learned and they won't achieve the results that were targeted. How can a person apply what they've learned if they don't remember what they've learned?

Given the causal nature of learning, we can't just measure one link in the chain and have a full picture of learning success. Indeed, measuring results by themselves tells us nothing about what went right or wrong in our learning designs and delivery.

Given all this, it's imperative to measure along the causal chain. We probably don't want to measure everything—it would require too many resources or too much time—but we can be strategic.

Backward chaining is one way to measure along the causal chain. We first select a result at LTEM Tier-7 or Tier-8. For example, if we're training sales managers to coach their sales reps, we might start with a Tier-8 goal of increased sales. We are training sales managers to coach their sales reps to be more effective, so an obvious way to measure their performance is their sales results.

We might select a Tier-7 goal to measure the quality of the coaching our sales managers are providing. Since their sales reps are being coached, those sales reps can provide good feedback about the quality of the coaching they have received from their sales managers.

Note that, in this example, we were able to find workable Tier-8 metrics to measure at a reasonable cost-benefit point, but this is not always the case. Many times, we will start our backward chain at Tier-7 by focusing on how well the learning transferred to work performance.

After selecting our Tier-8 and Tier-7 results metrics, we would continue our measurement along the causal chain by using LTEM as our guide. At Tier-6 we might measure the learners' success in a simulated coaching exercise. At Tier-5 we could ask learners to make scenario decisions illustrating key coaching decisions. At Tier 4A we could test learners' knowledge of the fundamental principles of coaching. At Tier-3A we could use a performance-focused learner

survey to measure the sales managers' perceptions of the effectiveness of the learning program.

Notice how our backward chain started with our ultimate goals and then targeted learning outcomes like Tier-6 task performance, Tier-5 decision making, Tier-4A foundational knowledge, and Tier-3A learner perceptions. It is these learning outcomes that give us the most revealing information about the workings of our learning programs.

A key advantage of backward chaining is that it focuses us on ultimate goals while also enabling us to have meaningful conversations with our key organizational stakeholders. Indeed, backward chaining helps us satisfy our stakeholders' concerns about the organizational relevance of our learning efforts. It also educates them about our key leverage points, potentially enabling us more grace and more resources to power our work.

8. Avoiding Scope Creep

The learning development process is rich in complexities; so is the learning-evaluation work we do. Indeed, when we do serious evaluation work—for example, doing more than surveying learners—there is sometimes pressure to go beyond our original project plan: to measure more, to add metrics, to fluctuate between different measurement goals. This scope creep causes all sorts of problems, including missed deadlines, overwork, resource deficits, higher costs, lower profits, harm to our reputation, and loss of trust. It's critical to avoid this scope creep.

LTEM can help by clarifying our measurement goals. *"It looks like we might be expanding our work beyond our original goals. Let's look back at our LTEM targets to see where we started. If we need to do more, that's fine, but we should expand our scope with purpose, not just slide into it. We ought to keep an eye on our time and costs as well. Is that okay with everyone: to look back at our original LTEM targets?"*

9. Pitching Evaluation

Whether we are building learning programs for internal or external customers, often we get pushback when we offer to do learning evaluation. Customers or sponsors don't see the value or don't want to pay for evaluation. This is highly problematic because the benefits of evaluation are huge. Rigorous learning

evaluation enables us to create maximally effective learning programs. Without evaluation, we are in the dark, and learning suffers.

LTEM, because it includes highly credible measurement targets—transfer to work performance, task performance, decision-making, and foundational knowledge—it lends credibility to our proposed measurement targets. Also, it paints a clear picture that learner perceptions—portrayed at LTEM Tier-3—are insufficient.

When pitching our stakeholders on evaluation, we can improve our persuasive success by avoiding go/no-go decisions. That is, we don't want to say, *"Here is our evaluation proposal. Should we include this in our project plan?"* Rather, we want our stakeholders to consider which evaluation options to consider. *"Thank you for brainstorming with us on how we might build a learning program for you. If we work together to build that program, we recommend these three evaluation measures to ensure we can gauge and report on our success. Which ones should we go with? Or shall we use all three to ensure we target learning transfer, decision-making, and learner perceptions about the effectiveness of the program?"*

10. Bolstering Credibility During Sales Discussions

Whether we are a vendor selling to our customers or are an internal learning team "selling" to our organizational sponsors, we need to appear credible in our sales discussions. LTEM can help by demonstrating the quality of our work and the rigor of our thinking.

"When we build a learning program, we aim to create Tier-7 transfer—that is, improved work performance: work performance that impacts key KPIs at Tier-8. We regularly measure our learning success at LTEM Tier-5 and Tier-6, because we believe decision-making and task performance are the best indicators of learning success we can measure during learning. Sometimes we measure Tier-4 knowledge when it's appropriate—when the knowledge is foundational to good work decisions—but we know that knowledge alone does not indicate success. In asking learners for their perspectives, we measure at Tier-3A, using performance-focused learner surveys, because the research is very clear that traditional smile sheets are not correlated with learning."

By indexing our learning goals with LTEM, we can demonstrate that we take our work seriously, that we utilize credible frameworks, and that we are dedicated to getting results.

Also, as the weaknesses of other evaluation frameworks become more well known, using LTEM lets us take less risk in being seen as stuck in the past using shopworn tools and antiquated thinking.

11. Uncovering Client Goals

Let's face it: Our clients, whether internal or external, aren't always clear about their goals for learning. Why should they be? They are not experts in learning, and that's okay. We can help by guiding them through a goal-clarification process. LTEM is excellent in this regard because it paints a clear picture of what learners might be asked to do.

We can put LTEM in front of them—ideally, the more conceptual illustration of LTEM, the one without all the details—and ask questions like these:

- If the learning program is successful, what outcomes would you expect to see (at each tier of LTEM)?
- Do you want learners to show that they can make better decisions based on what they learned?
- Do you want learners to demonstrate that they can actually carry out the kinds of realistic work tasks the training is targeting?
- Do you want the learning program to measure actual transfer—to determine if the learning produced improved work performance?

By getting answers to these questions, we can guide conversations around both learning evaluation and learning design. When we ask our clients whether they want us to measure actual transfer, we can negotiate what that would entail. When our clients tell us they want learners to demonstrate that they can perform tasks and make decisions, we can describe how traditional training focused on knowledge will not enable us to reach those goals—that more robust learning designs will be needed: for example, providing realistic practice spaced over time.

12. Game, Simulation, and Gamification Design

Training games, simulations, and gamification can support learning by increasing engagement and attention on key concepts and decisions. Unfortunately, they can easily be distracting and push people's attention toward less important considerations. LTEM can help by giving learning designers meaningful targets:

- Will my training game help learners gain foundational knowledge at LTEM Tier-4A?
- Will my simulation give people decision-making practice as targeted by Tier-5 work-realistic decisions?
- Will my gamified leaderboard compel my learners to practice work-realistic task performance at LTEM Tier-6?

13. Helping Us Choose the Right Vendors

Most learning teams hire vendors for various services and products. Vendors may have expertise or tools we don't have. They may enable us to scale. They may get things done more quickly or more efficiently than we can. There are many advantages to working with vendors.

Unfortunately, some vendors are mediocre, and some are downright incompetent in providing effective solutions. Worse yet, in the L&D industry there are thousands of vendors to choose from, making it extremely difficult to determine which are worthy and which should be avoided. Many a CLO has lamented hiring the wrong vendor—especially when a long-term contract keeps the circus in town for an extended stay.

LTEM can help. What do we want in a vendor? We want a vendor that provides effective solutions, and we want a vendor we can work with to achieve continuing improvements.

What enables effectiveness? What enables improvement? Effectiveness gets a boost when vendors rely on research-informed learning designs. That's a good starting point, but greater effectiveness can be achieved only by measuring learning factors that matter and then regularly using this feedback to create continuous cycles of improvement. In short, the best vendors are gathering meaningful data and using it to improve their learning programs and other learning solutions.

We can use LTEM to help us determine if prospective vendors have been using a rigorous learning-evaluation process. As we begin our discussions with a vendor, we might have a conversation like this: *"Let's use LTEM to index learning effectiveness. You've built lots of learning programs over the past year. For those programs, at which LTEM tiers have you been able to demonstrate learning effectiveness? For example, do you have evidence that your learners are able to*

demonstrate successful task performance or decision making? That is, can you show Tier-6 and Tier-5 results? If so, how do you measure at those tiers?"

We can also use LTEM to get a sense of how a vendor thinks about learning evaluation: *"Okay, brainstorm with us. If we build the learning program we've been discussing, what will you aim for our learners to be able to do to demonstrate that they've learned? Let's use LTEM as our guide. At what tiers will our learners be asked to demonstrate learning outcomes?"*

14. Guiding Our Vendors to Enable Full Learning Effectiveness

It's difficult to find the perfect vendor. Certainly now, most vendors are not doing as much as we'd like to enable effectiveness and cycles of improvement. That's okay, and it's not helpful to be nihilists and give up. What we want is a vendor who is willing to work with us in our efforts to measure and improve learning effectiveness.

Here again, we can use LTEM to guide ourselves and our vendors as we partner to produce maximally effective learning. *"For this project, we will require the following levels of measurement. We want to measure transfer to work performance (LTEM Tier-7). We want to use work-realistic scenario questions at Tier-5, and we'd like to use a performance-focused learner survey at Tier-3A. How might you conduct such measurements? And if you're not sure, are you willing to invest the time and resources in building up this capability, partnering with us to do this evaluation work, or partnering with a measurement consultant so we can get this done?"*

We can also guide vendors by putting our measurement requirements in our requests for proposals (RFPs). Within our RFPs we can have sections that prescribe such things as (a) evidence of previous learning-measurement practices, (b) required measurement targets on this project, and (c) a request for how the vendor plans to conduct their measurement work.

15. Learning Project Debriefs and Reporting

At the end of any learning-development project, we typically get together with our team and key stakeholders and debrief the project—determining what went well, what went poorly, and what we can do better next time. We also typically create a report or reports to describe our results, sharing these with important stakeholders. Such reporting closes the loop on our responsibility, provides us

with stakeholder feedback, and gives us an opportunity to educate stakeholders and demonstrate our capabilities.

LTEM can support these end-of-project events. In our reports, we can index our outcomes based on LTEM. *"After gathering pre-test and post-test responses on LTEM Tier-5 decision-making scenarios, we found that those who took the training in March improved their decision-making by 82% as measured in April compared with February. This compares to the group that will be trained later, which showed a non-significant 3% improvement from February to April. The trained group also rated the course highly on an LTEM Tier-3A performance-focused learner survey, showing the course had real strengths in motivating learners to apply what they've learned."*

Of course, we wouldn't just publish the learning outcomes; we would also add sections in our reports on lessons learned and plans for improvement.

Similarly, in our project debriefs, by using LTEM we can index our learning effectiveness on research-inspired categories of outcomes:

"Let's use LTEM to analyze how well we did in creating the outcomes we targeted. Our goal here is to gather lessons learned so we can work together productively on future projects. We know we can always do better, so the focus here is gathering lessons learned.

"Okay, let's start at LTEM Tier-6. Did our learners perform successfully when they were asked to demonstrate their skills on the relevant realistic tasks we used to measure their improvement? On Tier-5 did the learners improve on those scenario questions we validated? On Tier 3A, what did the learners report? Did our survey questions work, or can we make improvements?

"Again, let's focus on two things: (1) the learning effectiveness and what we can improve, and (2) were our evaluation metrics adequate in measuring our outcomes, or do we need to improve them?"

16. Role Modeling and Stakeholder Persuasion

LTEM can also be an education and persuasion tool. Simply by referencing LTEM and talking about how your learning designs will help your learners reach the higher levels of LTEM, we can pique the curiosity of our stakeholders, compel them toward improved learning designs, and get them comfortable with the idea that they can measure beyond learner surveys.

When they hear us talking about measuring LTEM Tier-7 transfer to work performance or building exercises that require learners to demonstrate task performance and decision making at LTEM Tier-6 and Tier-5, they'll see that they too can move beyond knowledge checks (Tier-4), learner surveys (Tier-3), and course completion rates (Tier-1).

LTEM's Design Rationale

In the previous sections, I outlined what LTEM is and how it can be used. Looking only at those discussions, you can already see how LTEM is useful. Still, you may find it helpful to understand why LTEM is so powerful and such an improvement over the learning-evaluation frameworks that went before.

In this section, I will describe why models and frameworks have value—what we hope they will do for us. I will outline the messages or nudges we hope are inherent in our learning-evaluation frameworks. And I will describe how the previously dominant model—the Kirkpatrick-Katzell Four-Level Model—has strengths and some major limitations in failing to nudge us toward good evaluation practices. Finally, I'll describe the messages that LTEM is designed to send. By examining all this, you will see how LTEM offers significant benefits. You'll also understand why we should abandon past models that send inferior messages.

Minds for Models

We humans build models, not only because it's fun and profitable, but because we need models to help us make sense of our complex world.[1] In truth, we need models because our minds are small. Oh, certainly, with exhaustive concentration we can discern even the most puzzling phenomena. But all of us (if truth be told anymore) are pre-inclined to simplicity—to the least energetic uses of our minds. We like things simple. Nothing wrong with that! It's natural. And it makes sense if you think about it. Who's got time to waste when we need to focus on what our boss wants, what show to binge watch, what outrage our political leaders will be caught emoting?

In the learning field, we build all kinds of models. We've got the four levels of Kirkpatrick, the Five Moments of Need, the five phases of ADDIE, the Nine Events of Instruction, the 4C/ID, the Decisive Dozen, the Learning Maximizer Model. We use these frameworks because we need them. They help us know what to do. But what if they didn't? What if they pushed us toward poor practices? Shouldn't we then discard them or fix them or replace them? Of course!

[1] I am using the term "model" in its colloquial form, meaning a framework or structure.

A model is only as good as the messages it sends and the nudges it emits. If the messages or nudges are helpful, it's a good model. If the messages or nudges are harmful, it could be time for a better model.

I'm using the term "messages" because it connotes a conscious process, and I'm using the term "nudges" because it connotes a subconscious process. Certainly there is overlap, and conscious and subconscious processing can both be in force. The key thing is to know that both messages and nudges impact our thinking and actions when we interact with models.

The bottom line—the thing I'm getting to in this report—is that LTEM is designed to send important messages and nudge our thoughts and actions in helpful ways. Later I will detail 30 messages and nudges that LTEM is designed to emit. For now, the most important thing to know about LTEM is that its messaging and nudging are based on wisdom distilled from the science of learning. LTEM has learning wisdom baked into it, which is helpful because the goal is to evaluate learning—or, more specifically, to evaluate how effectively our learning programs are enabling learning transfer.

This report will also describe the messages and nudges inherent in the Kirkpatrick-Katzell Four-Level Model. As the once-dominant evaluation framework in our field—one that is still used in many corners—we must understand its strengths and weaknesses. As I will describe, the Four-Level Model sends both helpful and harmful messages and nudges, but zero learning wisdom is inherent in its design.

Learning-Evaluation Messages and Nudges

In my book, *Performance-Focused Learner Surveys (the Second Edition)*, I introduced the concept of stealth messages to show how learner-feedback questions can send messages about factors that are critical to learning effectiveness. For example, a question about after-training support sends learners, trainers, and instructional designers a stealth message that after-training support is critical for learning transfer. A specific focus on management support can send a message that involving learners' managers is vital to learning application.

Stealth messages are messages we send that aren't perceived as being attempts at persuasion. They are potent for two reasons—first because they convey important ideas and second because their messages are more likely to be accepted as givens; less likely to be subjected to extended conscious deliberations in the minds of those receiving the messages. Stealth messages can be delivered with verbal communications, but they can also be delivered or strengthened using symbolic, visual, or metaphoric nudges. For example, the color red provides a stealth message of warning or danger in many human cultures.

Speed limit signs attempt to send conscious messages. The hope is that drivers notice them, think about their speed, and then slow down. Traffic nudging works more at the subconscious level. Traffic designers make roads narrower and put trees right next to roadways to slow drivers down. When drivers see that the road is narrow or see the trees, they are nudged subconsciously to slow down.

The ADDIE model, with the "E" standing for "evaluation," sends a stealth message that evaluation should be done at the end of the learning development process. ADDIE, then, uses a hybrid of messaging and nudging. The word "evaluation" sends a message. Its placement at the end of the model is also symbolic, visual, and metaphoric—and has become a nudge as much as a message.

I'm spending time on stealth messaging and nudging because they explain a great deal about the sorry state of our learning evaluation practices. Stealth messaging and nudging also provide us a lifeline—a way toward an infinitely more effective model of learning evaluation.

Here's the bottom line: If our models push us toward bad evaluation practices, we're likely to create poor learning evaluations. If our models push us toward good evaluation practices, we're likely to create effective learning evaluations. So, let us start by examining the messages we might want to send.

Important Messages and Nudges for Learning Evaluation

What messages should we send—if we can—when we consider learning evaluation? What messages and nudges should be built into our learning-evaluation frameworks? Here's my list of 30:

Learner Engagement and Activity

1. Just because learners ENGAGE IN LEARNING doesn't mean they will have learned. Therefore, measuring attendance is an inadequate way of evaluating learning.
2. Just because learners COMPLETE A LEARNING EVENT doesn't mean they learned. Therefore, measuring course completion is an inadequate way of evaluating learning.
3. Just because learners PAY ATTENTION doesn't mean they learned. Therefore, measuring learner attention is an inadequate way of evaluating learning.
4. Just because learners SHOW INTEREST during learning doesn't mean they learned. Therefore, measuring interest is an inadequate way of evaluating learning.

5. Just because learners ACTIVELY PARTICIPATE in learning doesn't mean they learned. Therefore, measuring participation is an inadequate way of evaluating learning.

Learner Perceptions

6. Just because learners say they LIKE A LEARNING EVENT doesn't mean they learned. Therefore, surveying learners on their general satisfaction—and on other factors not related to learning effectiveness—is an inadequate way of evaluating learning.

7. Surveying learners on FACTORS RELATED TO LEARNING EFFECTIVENESS is better than surveying them on other factors—those not related to learning effectiveness. It's better to query learners about their level of comprehension of key concepts, their motivation to apply what they've learned, the amount of relevant practice they received, and the after-learning supports they are likely to get than it is to survey them about their overall satisfaction or their willingness to recommend a course to other employees.

8. Just because learners REPORT THEY HAVE EXPERIENCED EFFECTIVE LEARNING METHODS doesn't guarantee they learned. Therefore, surveying learners on their experience during learning, while it can provide some useful information related to learning design, must be augmented with objective measures of learning.

Knowledge Comprehension

9. Just because learners can RECITE FACTS AND TERMINOLOGY doesn't necessarily mean they know what to do.

10. Some knowledge is TRIVIAL. It may be interesting and even relevant, but if it is not needed to help people make work-relevant decisions, we should not measure it to validate our learning success.

11. Some knowledge is FOUNDATIONAL. Comprehending foundational concepts helps people make work-relevant decisions, so measuring foundational knowledge can help us validate our learning success.

12. Just because learners can DEMONSTRATE FOUNDATIONAL KNOWLEDGE during or right after learning doesn't mean they'll be able to REMEMBER that knowledge later.

13. Measuring knowledge AFTER A DELAY is better than measuring it during learning or immediately after learning—when we are measuring for the purpose of validating our learning success.

14. Just because learners can show they COMPREHEND FOUNDATIONAL CONCEPTS doesn't necessarily mean they can FULLY TRANSLATE their knowledge into work-relevant decisions. They may be lacking other

knowledge needed to make good decisions. They may not fully understand how the work context might influence what is appropriate. They may need additional practice, etc.

Learner Decision Making

15. Measuring DECISION MAKING IS BETTER than measuring knowledge recitation or retention. Measuring decision making is also better than gauging learner perceptions.

16. Just because learners can MAKE WORK-REALISTIC DECISIONS during a learning event doesn't mean they'll remember how to make those same decisions later. Therefore, measuring decision making during or soon after a learning event is an inadequate way of evaluating learning.

17. Just because learners can MAKE WORK-REALISTIC DECISIONS doesn't mean they can perform successfully; they also must be fluent, appropriate, and timely in the way they act on those decisions. Therefore, measuring for task performance (that is, measuring decision making in action) is better than measuring for decision making alone.

Learner Task Performance

18. Measuring WORK-REALISTIC TASK PERFORMANCE is generally better than measuring decision making, because task performance is more authentic to work performance than decision making alone—specifically, it requires a fuller accounting of contextual cues and demonstrations of fluency in action. Note that task performance requires decision making, so we shouldn't ignore decision making when we are measuring task performance.

19. Just because learners CAN PERFORM TASKS during a learning event doesn't mean they'll REMEMBER how to carry out these tasks later. Therefore, measuring task performance during or soon after a learning event is an inadequate way of evaluating learning. Ideally, we'd like to measure task performance some time after learning.

20. Just because learners CAN PERFORM TASKS during a learning event doesn't mean they'll be ABLE TO CARRY THEM OUT in their work. They may lack motivation, hit obstacles, become distracted, or be persuaded not to utilize what they've learned. Therefore, measuring task performance during learning should not be considered a proxy for work performance.

Transfer to Work Performance

21. Measuring TRANSFER TO WORK PERFORMANCE aligns with the central goal of learning. When we provide learning, we hope learners can use what they've learned in their work.

22. In MEASURING TRANSFER, we must utilize RESEARCH-LIKE METHODS to help us determine whether it was our learning intervention rather than other factors that enabled the work performance.

Measuring the Effects of Transfer

23. In MEASURING THE EFFECTS OF TRANSFER, we should first establish that transfer occurred, while simultaneously utilizing research-like methods to help us determine whether our learning intervention or other factors enabled the transfer effects.

Measuring Work Results (Including Transfer and Transfer Effects)

24. It is helpful to consider separately TRANSFER and the EFFECTS OF TRANSFER. Whereas measuring transfer enables us to learn whether our learning program improved work performance, measuring the effects of transfer may tell us whether that work performance created beneficial effects.

25. MEASURING *ONLY* WORK RESULTS—whether we are measuring transfer and/or the effects of transfer—IS INSUFFICIENT AND DANGEROUS because we will be left blind to factors that enabled and/or made it more difficult to achieve those ultimate goals.[2]

26. It's important to consider all possible results of our learning interventions, WHETHER POSITIVE, NEUTRAL, OR NEGATIVE—avoiding bias, preconceived ideas, and dishonest manipulations.

27. It's important to consider the FULL SET OF STAKEHOLDERS that may be impacted by our learning interventions. We don't have to measure the impact on each stakeholder group, since not all will be central to our goals or responsibilities. However, we ought to at least consider whether we should account for various stakeholder groups such as beneficiaries, learners, the organization, coworkers/families/friends, customers, sponsors, investors, community, society, and the environs.

[2] So, for example, if our ultimate goal is to improve customer satisfaction, it's not enough to measure how learning impacted that satisfaction. We also need to know whether the learners could make good decisions, whether they could remember what they learned, and whether they comprehended what they learned in the first place. To see why this is essential, suppose we did our homework and determined that customer service is worse than it should be because employees have a knowledge problem—one that can be helped with an effective learning intervention. Suppose we train the employees, but customer service does not improve. If we haven't evaluated decision making, we won't know whether our training was deficient in helping learners make good customer service decisions. If we haven't evaluated remembering, we won't know whether our training was deficient in supporting remembering. If we haven't evaluated comprehension, we won't know if our training was deficient in supporting comprehension. Given that there is a causal chain from comprehension and decision making to remembering to on-the-job application to customer service, we must evaluate the full causal chain to know what happened.

28. USING SUBJECTIVE METRICS such as surveys, interviews, and focus groups has benefits but can also introduce biases, ignorance, faulty reasoning, and other cognitive deficiencies into our data. Subjective metrics must first be well designed. They must also focus respondents on constructs for which they have pertinent knowledge. Subjective metrics should be augmented with objective measures. Multiple sets of well-designed subjective metrics are better than one set of well-designed subjective metrics.

Learning-Design Encouragements

29. There are a NUMBER OF GOALS WE SHOULD HAVE as learning designers, including supporting our learners in building comprehension, remembering, decision-making competence, task competence, and perseverance in applying what they've learned to their job or other performance situations.

30. It is NOT SUFFICIENT TO SUPPORT LEARNERS ONLY IN COMPREHENSION, which is what too often passes for learning. At a minimum, we also need to support learners in remembering and in decision making, and ideally in task performance as well.

A good evaluation model will send messages and nudge thinking on all these issues—or at least hint at most of them. LTEM was designed with these core ideas in mind.

Let's now look at the messages sent in the Kirkpatrick-Katzell Four-Level Model.

The Kirkpatrick-Katzell Model's Successes and Failures in Messaging

Because the Kirkpatrick-Katzell Four-Level Model has dominated the practice of learning evaluation for the past several decades, it's important to examine its strengths and weaknesses as we consider LTEM's new approach.

The four-level Kirkpatrick-Katzell model sends many important messages.[3] It tells us we should focus not only on learning, but on creating on-the-job behavior and organizational results. It hints that learner opinions are not sufficient and should be a lower priority than the other three levels. It tells us the four levels are interconnected. These are critically important ideas—especially the concept that our end goal is not learning; that learning is, instead, a means to an end. We owe a debt of gratitude to Donald Kirkpatrick and Raymond Katzell for the stealth message that learning is not our end goal.[4] This is one of the most important ideas ever conveyed in the workplace

[3] Raymond Katzell originated the idea of separating learning evaluation into four parts (Kirkpatrick, 1956). Donald Kirkpatrick enhanced Katzell's formulation by creating labels for each of the four levels. He also, for several decades, tenaciously popularized the use of a four-part model.

[4] It will certainly surprise many that Raymond Katzell formulated the four-part notion of learning evaluation. You can read more here: https://wp.me/p9hau4-1bj, which can be found on my blog: https://www.worklearning.com/wills-blog/)

learning field. The Kirkpatrick-Katzell Four-Level Model was instrumental in getting the industry to accept this truth.

Unfortunately, the Four-Level Model also conveys—or is often understood as sending—other messages that are problematic. It sends the message that learner feedback is related to learning results, job performance, and organizational results; however, research shows that this is not true. Traditional smile sheets have been shown to be virtually uncorrelated with learning results and work performance.[5]

The Kirkpatrick-Katzell model also supports the connotation that measuring Level 3 (job performance) and Level 4 (organizational results) is more important than measuring Level 2 (learning results)—even though we learning professionals have much more leverage over learning results than we do over job performance and organizational outcomes. This idea is especially problematic when we interpret the model to mean we ought to measure *only* Level 3 job performance or Level 4 organizational results. It's problematic because it ignores the causal pathway starting at learning and moving to job performance and then organizational results. If we measure only Level 3 and Level 4 and get poor results, we will have no idea how our learning design contributed to these failures.[6]

The Four-Level Model is centered on training—hiding the role that prompting mechanisms (job aids, performance support, etc.) and on-the-job learning might play in learning and performance. If what gets measured gets managed, then we're not likely to even consider prompting mechanisms or on-the-job learning as targets for our evaluations. Indeed, until recently these non-training modalities were largely ignored in the workplace learning industry. Note that LTEM too does not encourage a focus on prompting mechanisms.

The Kirkpatrick-Katzell Model ignores the causal chain from learning to remembering to performance to results. More specifically, it has absolutely zero messaging about the critical importance of remembering—or that, while learning interventions can be good at supporting comprehension, they can be poor at minimizing forgetting. Ideally, we need a model that pushes us to use learning designs that support long-term remembering.

---

[5] Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997; Sitzmann, Brown, Casper, Ely, & Zimmerman, 2008.

[6] I am somewhat intrigued by Roy Pollock's idea (personal communication November 2017) that maybe we should measure work results routinely, and look at other learning results only if something goes wrong. I'm still uncomfortable with this, as I think we need more feedback (not less) on our everyday performance as learning professionals. But maybe there is a compromise that might work—perhaps a performance-focused learner feedback survey and a valid measure of work performance as Roy suggests.

The Kirkpatrick-Katzell model pushes us to focus on weighing outcomes, while it is largely silent about measuring the various learning supports critical to good learning outcomes. Most pointedly, the model ignores the role non-learning professionals—especially managers—can play in supporting the learning-to-performance cycle.

The Four-Level Model provides no leverage points for learning-design improvements. Indeed, with its focus on learner perceptions as a causal factor in learning results, it has distracted us from learning factors that really matter. Ideally, we'd want our learning evaluation model to encourage thoughtful reflection on how to improve our learning designs—pointing us to the importance of remembering and decision making, for example.

The Kirkpatrick-Katzell Model also fails to warn us about some of the worst evaluation practices: common practices like giving learning interventions recognition of achievement when learners attend, pay attention, complete, or participate in learning. For too long, many in the learning profession have used these irrelevant metrics as indicators of learning. A good learning evaluation model would disavow these ineffectual evaluation practices.

No model is ever perfect, and—given the gargantuan complexity of human learning—no learning evaluation model is likely to contain all the messaging we might want. In LTEM, The Learning-Transfer Evaluation Model, many of the above limitations are overcome, but some are left for us to consider outside the parameters of the model. Raymond Katzell and Donald Kirkpatrick, working in the 1950s, were not privy to today's advances in the science of learning, nor could they have foreseen the evaluation practices that would be popular in the next century. They built their model in response to practices of the 1950s. The key insights baked into the Kirkpatrick-Katzell model must be retained in any new model, inappropriate messages must be removed, and critical missing messages must be added where possible.

The following lists summarize the benefits and harms of the Kirkpatrick-Katzell model:[7]

Beneficial Messaging Inherent in the Kirkpatrick-Katzell Four-Level Model:

1. Tells us we should not just focus on learning, but we should also aim for on-the-job performance and organizational results.
2. Implies that learning is instrumental in achieving improved behavior and results.
3. Hints that learner opinions of the learning are not sufficient on their own and should be viewed as a lower priority than the other outcome measures.

[7] The Kirkpatrick New World Model introduces some minor tweaks, but generally conveys the same messaging as the original. Indeed, with its recommendation to think first about results and then work backward, it's not so much an evaluation model as a learning development model.

4. Implies that learner opinions, learning results, on-the-job behavior, and organizational results are interdependent.[8]

Harmful Messaging Inherent in the Kirkpatrick-Katzell Four-Level Model:

1. It fails to disabuse us of ineffective learning evaluation practices such as giving credit for attendance, completion, attention, and participation.
2. It ignores the causal chain from learning to remembering to performance and results.
3. It ignores the critical importance of supporting remembering and minimizing forgetting in our learning designs.
4. It ignores and fails to highlight the advanced learning goals of decision making and task performance.
5. It prompts us to value learner ratings as predictive of learning, work performance, and organizational results—something proven to be false as typically practiced.[9]
6. It implies that Level 3 and Level 4 (job performance and organizational results) are more important than measuring learning—even though we, as learning professionals, have greater leverage over learning.
7. It is also too easily interpreted to mean we don't need to measure learning results because job performance and organizational results are more important—thus ignoring the fact that learning is on the causal pathway to job performance and organizational results.
8. It pushes us to focus on learning outcomes and ignores the measurement of learning supports, including such things as situational cues, resources and time, cultural support, and management support.
9. The model is training centric, ignoring prompting mechanisms (like job aids and performance support) and on-the-job learning.
10. It conveys the idea that the only goal of learning should be organizational results, ignoring not only learning outcomes that directly affect learners but also other potential beneficiaries and injured parties.

[8] This is mostly a beneficial message except that learner perceptions have been shown *in practice* to be virtually uncorrelated with learning results and on-the-job behavior. To be specific, the beneficial message is that Level 2 (learning results), Level 3 (on-the-job behavior), and Level 4 (organizational results) are interrelated. On the other hand, perhaps new methods—like the performance-focused learner feedback approach—may enable a reliable positive relationship between learner perceptions and the other learning outcomes.

[9] But see the caveat in the previous footnote.

Limitations of Other Learning-Evaluation Frameworks

The Four-Level Model is not the only learning-evaluation framework with significant limitations. I have focused on the Four-Level Model because it has dominated our thinking for decades. We need to wean ourselves from it. Still, it's worth examining other major evaluation frameworks to understand how their limitations impact their usefulness.

The Phillips ROI framework essentially adds a fifth level to the Four-Level Model, along with recommending a process model for how we should evaluate; however, it is hamstrung by the same design as the Four-Level Model. It too smashes all learning metrics into one bucket. Therefore, it too has no learning wisdom baked into it.

The Brinkerhoff Case Method encourages evaluators to look at four areas of focus: (1) Learning Goals (the knowledge, skills, and attitudes that the training is designed to enable in the learner); (2) Critical Actions and their Contexts (the actions targeted by the learning program); (3) Performance Outcomes (the outcomes and achievements people will achieve as a result of the learning); and (4) Business Goals (the unit and business goals toward which the performance outcomes will contribute). Each is measured using learner surveys and interviews, which unfortunately are not good at measuring the target areas of knowledge and skills. Certainly, the four focus areas are critical and the methodology is sound. Given the methodology, however, the causal chain that runs from comprehension, remembering, decision making, and task performance to work performance is left poorly examined. There is not enough learning wisdom baked into the method to help us gain insights about our key leverage points.

The TDRp (Talent Development Reporting Principles) categorizes evaluation metrics into efficiency measures, effectiveness measures, and outcome measures. Unfortunately, this unwieldy mishmash of over 700 measures is overly complex, incoherent, unworkable, and not aligned with the science of learning. It cannot be recommended if our goal is to bring leverageable learning wisdom into consideration.

Let me add one caution here. In being critical of these evaluation frameworks, I don't mean to imply that there aren't things we can learn by examining them or using tactics inspired by them. We are all learning how to do better learning evaluations. None of us is all-knowing. No model or framework is perfect. We are coming out of the dark ages in learning evaluation, so we must be open to learning from each other.

Details about LTEM

LTEM, The Learning-Transfer Evaluation Model (pronounced "L-tem" and written as "LTEM"), is designed specifically to help organizations and learning professionals determine whether their evaluation methods are effective in providing valid feedback. Previous evaluation models have not done this sufficiently.

LTEM is composed of eight tiers[10]—starting at the top, with the effects of learning transfer, and moving down the page toward methods of learning evaluation that should not be used to validate our learning success.[11]

The model is made available in a single-page, tabular format to ease comprehension. It is annotated and color coded to enhance the clarity of its messaging. The color coding utilizes red to indicate poor or inadequate methods, yellow to indicate mediocre methods, and green to indicate good or useful evaluation methods.[12] The color signals are meant to represent what each tier can tell us about our success in creating learning transfer. They do not signify the value of the tiers for formative evaluation or for other goals we might have.

The model is designed to be relevant for all learning interventions, including classroom learning, elearning, mobile learning, on-the-job learning, self-study learning, etc. Where learning occurs, we can evaluate it. We may have to change terminology or modify our evaluation practices, but the new model is designed to help bring wisdom to our efforts.

The model, while an exponential improvement over the Kirkpatrick-Katzell Four-Level Model, still leaves some important messages unstated. For example, it does not focus on evaluating prompting mechanisms like job aids or performance support. This is intentional. When models get too complicated, people don't use them.

Each of the tiers of the model will now be described, including the rationale and reasoning behind them.

Tier 1—Attendance/Completion

Believe it or not, many organizations will certify learning by checking attendance or completion rates. This is a terrible idea because neither attendance nor completion guarantees that learning occurred. Learners have been given credit for signing up for courses, starting courses, sitting through learning events, and completing learning experiences. None of these attendance metrics is acceptable as a measure of learning.

---

[10] I use the word "Tiers" instead of "Levels" to distinguish LTEM from the Kirkpatrick-Katzell Four-Level Model. This change was made in the 12th version of the LTEM model because, in practice, I found that people wanted to discuss LTEM in comparison to the Four-Level Model. Saying "Level 4" could get people thinking about the Kirkpatrick-Katzell level four when the intention was to talk about LTEM level four—or vice versa.

[11] One subtlety here is that, just because a tier represents metrics that should be avoided in validating our success, we might still use those metrics to our benefit as a formative evaluation practice—that is, in trying to understand more about how our learning program is working. So, for example, if we claimed our program had success when we saw that 98% of our learners engaged actively in discussion groups, we might be mistaken in our "success" conclusions. Perhaps our learners reinforced the wrong ideas in those discussions. Maybe the learners failed to grasp the key concepts, couldn't remember them two days later, or were demotivated by our program.

[12] The color coding is aligned with some cultural expectations but misaligned with others. So, for example, red can connote bad or good depending on one's experiences or cultural influences. Color expectations are too varied to work universally, so I've left the colors as originally designed, probably reflecting my lifetime of experience living in the United States.

People may attend or complete a course but learn the wrong information, or end up with misconceptions, or not believe in the main points, or be unmotivated to put what they learned into practice. By seeing attendance or completion as success, we are blind to these failures.

Tier 2—Learner Activity

Too often, we use learner activity as a gauge of learning. We do this both formally and informally. For example, we may formally measure the number of words a learner uses in an online forum. We may informally gauge learning success by the "energy in the room" when learners are involved in small-group discussions.

There are three main ways we might measure learner activity: attention, interest, and participation. All three are unacceptable as measures of learning because our learners may pay attention, show interest, and participate actively—yet not learn what they needed to learn.

What are some examples? Some organizations give credit in elearning programs when learners are periodically presented with a button to click to show they are paying attention. Other elearning programs measure attention by monitoring timing of mouse clicks. Measuring attention is an unacceptable metric because attention does not guarantee learning.

Instructors, whether online or in a classroom, often give credit for class participation. Sometimes peers are asked to evaluate each other on various activity criteria. Rarely, but sometimes, outside observers evaluate learner activities. Note that activity metrics go beyond mere attendance; instead, they attempt to gauge such things as how active, engaged, helpful, or cooperative a learner appears to be during learning.

While learner activity may seem to be indicative of good learning design, it is inadequate because activity does not demonstrate learning results. Learners may fail to learn. They may learn the wrong things. They may learn things poorly. They may focus on low-priority ideas instead of critical constructs. They may fail to be prepared to use what they've learned. Some activities are useful. Some are a waste of time—even though they seem useful. Some activities are harmful.

With today's digital technologies, it is tempting to count learner activity by using objective metrics (for example, the number of words written on a discussion board, number of social-media posts, or quality of written concepts as gauged by artificial intelligence). While these measures may remove the subjective element, they still suffer from the same problem as other activity metrics. People can be objectively active in learning, yet still fail to learn, or learn poorly, or learn the wrong things, or be unmotivated.

Tier 3—Learner Perceptions

Learners are often queried about their perspectives on the learning experience. We do this both formally and informally. We provide learners with formal surveys (aka smile sheets, happy sheets, feedback forms, etc.). We also might ask a classroom of learners to share what they think about the learning they've just completed. Even more informally, we might wait to hear what the "word on the street" is about the learning—either from the learners themselves, the learners' managers, or even from other employees.

Most commonly, we formally survey learners about their satisfaction—about whether they'd recommend the course to others, whether they felt the instructors were competent, whether the course met their expectations, etc. Unfortunately, such traditional smile-sheet questions have been found to be virtually uncorrelated with learning results.[13] While such surveys are useful to gauge the reputation of the learning experience, these smile sheets cannot be recommended as a measure of learning.

Learners can be asked questions that more fully gauge learning effectiveness. In my book, *Performance-Focused Learner Surveys*, I detail "distinctive questions" that can be asked—questions that focus on research-based factors such as whether the learning supported comprehension, remembering, and motivation to apply what was learned, and supports for after-learning follow-through.

Despite these improvements in survey methods, surveys focused on learning effectiveness create proxies of the constructs they target. For example, surveying to determine the level of realistic practice learners receive is just a proxy for whether the learning is supporting long-term remembering. Ideally, we'd measure remembering directly—as encouraged in the higher tiers of LTEM. Surveying for effectiveness is designed based on learning science findings. More research is needed to verify that surveying for effectiveness is related to the targeted constructs in the many contexts where learner feedback is gathered. As of now, it seems fair to say that surveying for effectiveness may provide adequate data—at least for further hypothesis testing—but it should be periodically augmented with objective measures of learning effectiveness.

Tier 4—Knowledge

Trivial vs. Foundational Knowledge. We can test learners on low-importance knowledge or high-importance knowledge. LTEM distinguishes between "trivial knowledge" and "foundational knowledge." Where foundational knowledge is critical to work decisions, trivial knowledge is information that is not critical to work decisions.[14]

---

[13] Several meta-analyses have found a lack of correlation between traditional smile-sheet ratings and learning. Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997; Sitzmann, Brown, Casper, Ely, & Zimmerman, 2008; Uttl, White, Gonzalez, 2017.

[14] This distinction was introduced in LTEM Version 13.

Let me use an example from the learning field to add some clarity. If we are teaching new instructional designers, we want them to know that realistic practice is critical in helping learners remember. It is foundational knowledge. We do not need them to memorize the fact that realistic practice is aligned with the research-informed recommendations of Robert Bjork, the famed learning researcher. This information is not required for them to make work decisions. It is trivial knowledge.

The distinction between foundational and trivial knowledge is critical because too often in the past we would evaluate our learning programs by asking learners questions about unimportant facts and terminology.

Of course, there are some situations when facts and terminology do aid work decision making—in those rare instances where it's a person's job to recall facts and terminology. For example, supermarket cashiers who need to memorize product codes might be properly evaluated using their ability to remember product-code knowledge.

Measuring for Comprehension or Remembering. Knowledge is often tested to ensure learners have understood the concepts, but previously we gave too much credence to short-term demonstrations of knowledge. Knowledge can be tested soon after learning or after a substantial delay—after several days or so.[15] This dichotomy is critical because remembering is critical to most real-world learning. When learners are tested soon after learning, we cannot be sure they will remember what they've learned. We cannot be sure that, when they encounter relevant work situations, they will remember the concepts they would benefit from remembering.

We can call near-term tests of knowledge "knowledge recitation" or we can say we are testing for comprehension or understanding. When learners are tested after a substantial delay—say, three days or more—we can say we are measuring *knowledge retention*—that we are determining how successfully the knowledge has been remembered.[16]

[15] Over the years, as I studied the research on learning, memory, and instruction, I noticed that many learning experiments distinguishing between short and long retention intervals used week-long delays as the longer interval—and, more importantly, that relatively long delays didn't show much differentiation after a week or more. Certainly, people forgot more over time, but the experimental findings don't tend to differ significantly for, say, a one-week delay versus a two-week delay (when both are compared to a one-day delay, for example). To be practical, I've set longer durations at "several days or more"—meaning three or four days or more—because it's my assumption that many workplace learning stakeholders would balk at routinely using week-long or longer delays in their learning designs and/or evaluations.

[16] There is nothing wrong with measuring short-term recall as a formative evaluation tool. Indeed, one of our goals as learning designers is to make sure learners comprehend clearly. However, when we attempt to validate our success by using short-term measures, we can be blind to whether our learning program has enabled our learners to remember.

When we test learners' knowledge during or right after they learn, we are only testing whether they can retrieve information that is highly accessible in long-term memory. There are two reasons that information tested under these conditions is highly retrievable from memory: Learners have not had time to forget, and they are likely still in the learning context where stimuli are continuing to trigger memory retrieval. Such immediate tests are not authentic because, in most work situations, the information learned will not be highly accessible in memory. It will have faded to lower levels of accessibility.

When we measure trivial knowledge during learning or right afterward—a too-common practice—we experience two crippling learning-evaluation problems:

- Knowing unimportant facts and terminology will not enable decision making and task performance. To put it simply, trivial knowledge does not typically enable performance.
- Retrieving information so soon after learning is an untenable metric because immediate recitation does not guarantee the ability to retrieve information later. Our most common practice in measuring knowledge—trivial information measured soon after learning—is not acceptable.

Bloom's Taxonomy Is Not Transfer Focused. Knowledge, of course, is not all one thing. Some knowledge is composed of facts and trivia. Other knowledge is composed of more meaningful constructs, including concepts, principles, generalizations, processes, and procedures. Indeed, we could break down knowledge into myriad subcategories. A common method of categorizing knowledge is Bloom's Taxonomy, originally published in 1956. But the various updates of the taxonomy in the 2000s show the multitude of ways that knowledge can be subdivided.[17] Interestingly, Benjamin Bloom and his colleagues created their taxonomy based on actual evaluation tools that were used in schools and universities—tests using essays, completions, short answers, multiple choice questions, etc.

None of the Bloom's Taxonomies—not Bloom's, not Marzano's, not Anderson and Krathwohl's—is useful for our purposes. Yes, you read that correctly. First, the many variants of Bloom's Taxonomy are too complicated. Second, and more importantly, they don't focus on the future utilization of learning. That is, they don't focus on work transfer. They focus on testing of the kind done in schools. By keeping a laser-like focus on transfer, we are better positioned to have an evaluation model that is practical and relevant to our targeted outcomes.

While we don't have to overcomplicate our evaluation of knowledge, we should focus on the most meaningful aspects of the knowledge that is targeted for learning. Instead

[17] Bloom, 1956 (or Bloom, Engelhart, Furst, Hill, Krathwohl, 1956), Marzano, 2001, Marzano & Kendall, 2007; Anderson & Krathwohl, 2001.

of focusing on definitions and terminology, we should focus on the concepts, principles, and wisdom that relate most directly to the learners' future use of the information. We should also keep in mind that we can evaluate procedural knowledge in the decision making and task performance tiers of the LTEM framework—Tier-5 Decision Making and Tier-6 Task Performance.

An Extra Note on Realistic Remembering

I'm making an intentional dichotomy between short-term remembering and relatively long-term remembering. This distinction is one of the most critical concepts in learning. If we prepared our learners only for the short term, much of their training would be wasted. They simply wouldn't remember what they learned and, when they needed the information in their work, they wouldn't retrieve it from memory. Too much training is built just like that—creating only short-term effects—which is another reason this dichotomy is so essential.

You may wonder if this applies when we teach people and they begin to apply what they've learned the very next day or shortly thereafter. Let's say we have a one-day workshop and we teach people 20 key concepts. Let's say the learners take our workshop on Monday and begin to apply what they've learned on Tuesday. How many of those 20 learning nuggets are they likely to use that first day? Maybe 10. How many the rest of the week—on Wednesday, Thursday, and Friday? Maybe five more. Already we can see—even for this seemingly just-in-time learning intervention—that learners must remember 50% of what they learned longer than a day, and 25% of what they learned longer than a week. The conclusion from this thought experiment should be clear: Only rarely do learners immediately use everything they've learned—mostly when they learn only one or two things and utilize them right away. The truth of it is that, for almost all our learning interventions, our learners will have to remember a significant portion of what they learned.

The dichotomy between short- and long-term remembering plays out in three places in the LTEM model: in Tier-4 Knowledge, Tier-5 Decision Making, and Tier-6 Task Performance. In previous versions of LTEM, I used the terms "knowledge recitation" and "knowledge retention" to make the distinction in Tier-4. In Tier-5 I represented the distinction by using the terms "decision making" and "remembered decision making." In Tier-6, we could use the terms "task performance" in comparison with "remembered task performance." In each case, the remembered measure is closer to the time lags learners will experience in their real work situations—when remembering is essential to their performance.

To be practical and specific about this, testing for short-term effects should involve measuring within the same day in which learning took place (i.e., within about eight hours). Long-term remembering can be operationalized to involve delays of three days

or more after learning, to be practical—although delays of a week or more are generally more realistic.

Tier 5—Decision Making Competence

In the four-level Kirkpatrick-Katzell model, Level 2 was designated as "Learning." Too often, "learning" was interpreted as "knowledge recitation"—asking learners to regurgitate facts and terminology they had just learned. But learning results are so much richer. At a minimum, learning results can constitute comprehension, the ability to make decisions, and the ability to successfully engage and complete realistic tasks.[18]

Decision making is clearly an essential result of any learning initiative that intends to support subsequent behavior. When we teach managers to be better supervisors, we hope they'll make better *decisions* when they are interacting with their direct reports. When we teach employees to avoid engaging in sexual harassment, we hope they will make better *decisions* about what to say and what to do. Even in purely educational settings, we should design our learning to help learners make better *decisions*. When we teach algebra, we hope our learners will be able to *decide* when to reduce like terms to one instance in an algebraic expression. When we teach history, we hope (or we should hope) our learners will be able to make better *decisions* when they enter the voting booth.

Training that is aimed only at creating awareness—without any expectation that behavior should change—is training that is likely a waste of time. Learners may become attuned to important concepts, but they are likely to (1) forget what they've learned (because of a lack of realistic practice), (2) be unprepared to use what they've learned (because cognitive links have never been made between real-world situations and the learned concepts), and (3) be insufficiently motivated to apply what they've learned (because they haven't considered the real-world implications).

At a minimum, we should design learning to support realistic decision making and we should evaluate our learning interventions based on how well they support learners in gaining decision making competence.

---

[18] Once, intending to lovingly augment the Kirkpatrick-Katzell Four-Level Model, I designated a Level 2.5 as learner decision making. Of course, decision making is not the only augmentation needed. I've already mentioned task performance, and perhaps there are other learning results we might consider as well. Note that, while we might consider adding "motivation to apply" as another learning result, there are good reasons to avoid making it a separate evaluation target. For those interested, I didn't add a separate tier to the new model focused on "motivation to apply" because motivation (1) is really hidden inside human cognition, (2) is thus not directly measurable, and (3) can be indirectly measured through learner surveys, which are already represented in the new model.

Given the imperative to account for short- and long-term remembering, we can measure decision-making competence in the short term (within the day it was learned) or in the long term (for example, after three or more days).

Measuring short-term realistic decision making is a giant leap from lower tiers, but it is still insufficient as a learning metric because it does not evaluate whether learners can maintain their newly-learned decision making ability over time—to support them in using their skills on the job.

Measuring long-term realistic decision making is an important benchmark for learning professionals. Indeed, it's a juicy target because (1) it's a relatively authentic demonstration of work competence and (2) measuring decision-making is a relatively moderate investment for the benefits expected. Having a signal in our evaluation model that nudges us to measure decision making is huge! It ensures we are measuring and designing for real work performance.

Examples of How to Measure Decision Making Competence

To evaluate realistic decision making, we need to present learners with realistic situations and prod them to make decisions that are similar to the types of decisions they will have to make on the job. There are two (or three) ways to do this:[19]

- We can present learners with realistic scenarios and ask them to make realistic decisions based on their interpretation of the scenario information.
- We can ask learners to make decisions in high-fidelity simulations.
- We can ask learners to make decisions in real-world situations—for example, on the job.[20]

Note that real-world decisions on the job will often push learners into the next evaluation tier—Tier-6 (Task Performance)—because, in the real world, we don't usually just make decisions; we usually make decisions and take actions together.

Scenarios can be text based or can involve video and audio. They can involve forced choice responding, as with multiple-choice questions. They can involve multiple-option responding. They can involve open-ended responding or some other simple response

[19] Here we are indebted to Sharon Shrock and Bill Coscarelli, whose evaluation taxonomy—introduced in their classic text, *Criterion-Referenced Test Development*—utilized three categories of realistic performance: (1) Real-World, (2) Simulation, and (3) Scenario-Based Decision Making.

[20] Note that, whereas scenario and simulation decision making are decidedly learning situations, if we ask our learners to make real-world decisions, this might seem to jump our evaluation up to Tier-7 in the model: Transfer. However, although there is some ambiguity in the case of learners performing in real-world situations, asking our learners to engage in real-world decision making for the purpose of learning has a different psychological character and can therefore be appropriately separated from regular on-the-job decision making (where learners have left the formal learning experience and have transferred their learning without our direct intervention).

method. Scenarios can be short and simple or they can be long (or threaded) as if in a story, or they can involve multiple layers of analysis before decisions are made—like in case studies.

Simulations go beyond scenarios by providing more realistic contexts and often more realistic responding as well. Note that there is not always a definitive line between scenario-based metrics and simulation-based metrics. Often, in practice, things called "simulations" are really fancied-up scenarios. It doesn't matter much for our purposes. The keys are:

- Whether learners are provided with realistic background contexts
- Whether they must decipher what the context means (without help or hints—because we're assessing here!)
- Whether they have a range of options from which to choose to indicate the decisions they are making
- Whether they are compelled to actually make a decision

Bridge between Decision Making and Task Performance

The distinction between decision making and task performance is complicated and important, so I'm going to provide a long discussion to introduce it.

Task performances comprise decision making (as exemplified in Tier-5) and task performance. Here is an example. Suppose we teach managers to give their direct reports performance feedback and then we assess their decision-making ability in knowing what to say and when to say it. If we do that, we're measuring their decision-making competence. But then, suppose we go further and assess their ability to voice the feedback (what tone of voice and body language to use). If we do that, we're measuring their ability to take action after they've made a decision. By measuring both decision making and action taking, we're measuring task competence.

When we teach, we often separate decision making and task performance. For example, if we're teaching electricians, we can give them practice troubleshooting circuit problems using a diagram that represents electric circuitry. This is decision-making practice. They are not taking action; that is, they are not finding real wiring to test, using tools in a realistic way, etc. Similarly, we can give salespeople practice deciding what to say when faced with customer objections without having them actually voice the words they would use or the tone of voice they would employ. We can give nurses practice deciding what to do when faced with simulated patient information, without having them actually take a person's blood pressure and other vitals. If we measure this decision-making ability alone, we are doing Tier-5 work. Again, Tier-5 measurements ask learners only to make decisions, not to implement those decisions.

Conversely, we can also give people practice on task performance without asking them to make decisions. We can say, "Okay—role play what you'd say to an angry customer who is upset at having to wait in line." In such a situation, we have essentially made the decision for the learners. They haven't had to decide that the customer is angry because of a long line. Similarly, we can train a nurse on how to listen through a stethoscope to the sounds of a person's airways without having them decide when to use such a procedure.[21]

To reiterate, we can focus on decision making, action taking, or both. In LTEM, when only decisions are measured, we're in Tier-5; when decisions and actions are measured as full tasks, we are in Tier-6.

Special Note to Introduce the SEDA Model

The SEDA Model helps us clarify the distinction between decision making and task performance. Originally developed to help learning designers write powerful scenario questions—and taught in my LTEM Boot Camp and in my workshop on writing scenario questions—the SEDA Model is a practical tool for multiple learning-design purposes. It works because it is based on a simple truth about human behavior.

Consider what people do in their real-world experiences. They are constantly faced with *Situations*. They *Evaluate* those situations—in other words, they make sense of them. Once they understand a situation, they make a *Decision* about it. Then they take an *Action*. This can be diagrammed as follows:

Situation → Evaluation → Decision → Action

SEDA is the acronym formed by the first letters of the words "Situation," "Evaluation," "Decision," and "Action." The SEDA Model, then, nicely simplifies what people do in the real world. Each action creates a new situation. Each new situation provides feedback on the previous action, decision, and evaluation.

When we design learning, the SEDA Model can be our guide. We know—from research on how context influences the learning-to-remembering process—that real-world situational cues trigger thinking and action. Human cognitive processing is reactive largely to triggers, whether those triggers come from a person's environmental context or from their internal working memory. SEDA, although a simplification of human cognitive processing, captures the basic elements needed to guide learning design. The key for learning is ensuring learners practice every aspect of the SEDA Model. We need

[21] I should note, for precision, that—when we attempt to determine what constitutes a decision and what constitutes an action—there will always be some subjectivity. In truth, different levels of abstraction are at play. For example, I labeled "a nurse using a stethoscope" as an action, but the nurse may have to decide how to hold the stethoscope, where to place it, etc. Don't worry; in practice, the distinction between decisions and actions is usually straightforward.

to place learners in realistic Situations (whether real-world situations, simulated situations, or scenario-based situations), we need to require our learners to Evaluate those situations, ensuring they make sense of those situations on their own, and we need to prompt learners to make Decisions based on their evaluations—and take Actions based on their decisions.

This sounds simple, but it's often easier to comprehend the value of SEDA by looking at failures in learning design. I will do this quickly so we can get back to our focus on evaluation.

I'll start with one of my own failures. I used to teach simulation-based workshops on leadership topics. I taught content, then put learners in elearning-based simulations where they made decisions based in realistic leadership scenarios. Let's examine this by using the SEDA Model. I presented my learners with Situations, which they Evaluated (made sense of) before making Decisions about what to do in those simulated situations. What they *never* did in my workshops was practice the Actions they would have to undertake to implement their decisions. They never practiced what words to say, what tone of voice to use, what body language to emanate. We might say the learning design was a SEDA (with a grayed-out "A" to signify that learners did not practice Actions).

What about an elearning-based simulation that teaches learners how to diagnose a medical episode? *"An older, middle-aged man enters the ER. He's an alcoholic and is complaining about lower abdominal pain."* Learners are asked what to do. Using SEDA, we would be presenting a Situation, but would NOT be giving learners a real chance to Evaluate that situation. Did you notice that the simulation told the learners the man was an alcoholic? The simulation basically evaluated the situation for the learner. We might grade the experience as a SeDA, using a lowercase "e" to signify a partial Evaluation and a grayed-out "A" to signify that no actual actions were practiced—that is, learners never had to actually examine a person's body, call in a prescription, etc.

What about a statistics class that teaches one topic a week for many weeks? So, for example, we might teach Chi-Square in the first week, t-tests in the second week, f-tests in third week, and so on. During the week, we give learners extensive practice calculating the statistics. At the end of each week, we give learners a test asking them to calculate the statistic taught that week. Using SEDA, we clearly have given learners plenty of practice in taking Actions—doing the math to calculate statistics. What we have *failed* to give them are any real-world Situations that need to be Evaluated. Nor have we given them practice in making Decisions about which statistic to use—and when. In the real world, people using statistics are faced with some sort of research effort to work through. They must Evaluate those Situations and Decide which statistic to use. We would have to grade this statistics class as SEDA, with a grayed-out "S," "E," and "D." Note how harmful this design is! Learners learn procedures for calculating

statistics—something they will *never* do in the real world because statistical programs do this grunt work. Moreover, they never actually practice the most important skill—figuring out which statistic to use given a variety of background scenarios depicting different research situations.

One more—and this one's the killer. What about lawyers learning the law or students in a marketing class learning marketing concepts or employees being onboarded by learning about the history and values of their new company? Let's say these learners are provided with scintillating speakers who cover concepts and highlight important principles. Using SEDA, we can see that the learners were *NEVER* presented with Situations—and, of course then, they never had to Evaluate situations, make Decisions, or take Actions. They got no practice in any of the SEDA components! An empty SEDA—a ghost SEDA—and, incidentally, a very poor learning design: one that leaves learners with zero skills.

I'm not arguing that every learning event must incorporate all four components of SEDA. While that may be ideal—and, while SEDA may help us in aiming for full learning effectiveness—we can also use SEDA to help us keep track of the SEDA components we are providing and those we are missing. For example, going back to my leadership simulation days, I might have used SEDA to see that I was missing the Action component. I could have then redesigned the course to include more action practice or I could have ensured learners had job aids or performance support or additional practice before they attempted to implement the decisions they had learned to make.

With that brief introduction to the SEDA Model, it should be clear why, at a minimum, we need separate tiers for Decision Making and Task Performance.[22] We want our evaluation model to prod us to create learning designs that are fully effective. By focusing on Decision Making in Tier-5, the new model promotes Evaluation and Decision Making in our learning designs. By focusing on Task Performance in Tier-6—and remember: Task Performance includes both decision making and action practice—LTEM promotes the use of all four components of SEDA. In other words, Tier-6 provides practice in all the real-world behaviors needed for our learners to be fully effective in their work or in other targeted performance situations.

[22] In the Learning-Transfer Evaluation Model, I'm including the Situation, Evaluation, and Decision components from SEDA within the Decision-Making Competence tier (Tier-5). I'm adding the SEDA Action component to create the Task Competence tier (Tier-6). Therefore, the Task Competence tier of the model includes all four SEDA components. It presents learners with Situations and has them Evaluate those situations, make Decisions about what to do, and take the Actions determined by those decisions.

Tier 6—Task Performance

We can now focus on Tier-6—Task Performance.

Let's start by looking at task performance when demonstrated soon after learning. If our learners can demonstrate their ability to perform a task right after learning, they are well on their way to transferring what they've learned. Demonstrating full task performance is a very strong learning result. But there is one catch: If learners demonstrate task performance soon after learning, we still don't know whether they can remember that competence over time. Measuring task performance during or soon after learning is not sufficient to enable us to certify learners as prepared for transfer.

On the other hand, if task performance can be demonstrated several days or more after learning, we can certify learners as task competent. Again, this doesn't mean they have transferred learning. They haven't. They've merely shown they can perform tasks while situated in a learning context. Actual transfer is represented in Tier-7.

Examples of How to Measure Task Performance

To measure task performance, we must provide learners with a full SEDA experience. We need to present them with realistic Situations, have them Evaluate those situations (again, without help or hints—because this is an assessment!), enable them to make Decisions, and have them take Actions in line with those decisions.

If we're teaching people to use PowerPoint® so they can make effective presentations to people they are training, we'd provide them with PowerPoint, some content to teach, some information about the learners, and some specific learning goals to achieve. We might then have them create a set of PowerPoint slides and have them make a training presentation to some simulated trainees (maybe their fellow learners, class instructors, or other learning professionals). In doing all this, we've presented them with a realistic situation (i.e., PowerPoint, training goals, learners) and had them Evaluate all the requirements, Decide on the design of the slides and the learning flow, and take Action in using what they've created.

If we're teaching supervisors how to run a meeting, we could evaluate their performance running a simulated meeting with actors. If this is too cumbersome or expensive, we could also use the SEDA Model to design alternatives. If we had to keep costs to the bare minimum, we could create scenario questions that described meeting situations and have our learners evaluate those situations and make decisions about what to do. We would have a SEDA result that provided Situations, Evaluation of those situations, and Decisions. What we wouldn't have is an assessment of how well people could take Action. So, we might then decide to add another evaluation element. We could develop short video clips of meetings with a perspective that makes the learner part of the meeting and requires them to voice their responses at certain times. Responses could be video recorded and evaluated based on criteria taught in the

learning. Ideally, we might use experts, including instructors, to evaluate these action performances; but, in some cases, it might be acceptable to use peer evaluations to save time and money.

As you can see, evaluating task performance can be much more demanding than measuring at lower tiers, but also more aligned with real-world competence.

Tier 7—Transfer to Work Performance

Learning transfer is defined by two criteria. First, people have had to previously engage in some sort of learning experience. Second, they must use what they've learned on the job (or in some other targeted performance situation).

Special Note on the "Targeted Performance Situation" Idea

It's easy for us to think about someone putting their learning into practice on the job. But people can put what they've learned into practice in other ways as well. A person who takes a leadership class may use what they've learned in helping to organize a local charity or run a youth soccer team. These transfer successes may not have been the goal of the original learning investment, but they do represent transfer nonetheless.[23] For this reason, I've tried to use the word "work" instead of the word "job" in the optimistic hope that "work" can be broadened to include other endeavors besides on-the-job activities. Aren't we doing work in calculating an algebra problem, in practicing a foreign language, in learning to bake bread?

Not all learning investments are designed with job performance as the goal. Education settings are the clearest example of this. College students may learn algebra so they can learn trigonometry and calculus. High school students may learn history and social studies so they can make better life choices or vote for better elected officials.[24] So, for an algebra class, the targeted transfer situation might be the students' future trigonometry class. For history, one targeted transfer situation might be the voting booth; others might include future history classes, discussions with friends related to history or politics, or reading historical fiction.

I've developed the Learning-Transfer Evaluation Model to work for both workplace learning and education settings. One warning here: Because my career has focused mostly on workplace learning, I may have unintentionally framed things more readily for workplace learning. Still, as I worked on the model—and this report—I did keep one eye open to the world of education. I'm confident the model has relevance for education as well as workplace learning. Indeed, some educators have told me they have found

[23] We should probably separate transfer that was targeted in our learning design from transfer that is incidental or a lower priority. We typically invest in learning to gain some sort of benefit. We can certainly have multiple goals, but we might want to be clear about what our bottom line is for successful transfer before we design and develop our learning initiatives.

[24] Insert your own joke here.

benefit in using LTEM. Let me add that I think students would benefit if more teachers and professors kept their learners' future performance situations in mind.

By designing for transfer, we create learning that is more likely to be remembered and utilized. What's the value of learning something if that knowledge is locked in our brains, never used—fading into fuzziness over time? Educators who want to use the model may have to occasionally reframe some of the words or the emphases, but the model is intended to have the same benefits for education as it does for workplace learning and performance.

When we measure for transfer, then, we must select a relevant performance situation to target. We don't willy-nilly want learners to use what they've learned. We really hope they use what they've learned in certain situations or contexts. When we provide leadership training, we hope supervisors will use what they've learned in situations in which they are interacting with their direct reports. When we teach PowerPoint to help people craft more effective presentations, we hope our learners will use what they've learned while they are using PowerPoint to make presentations.

It might be helpful to consider two levels of transfer: assisted transfer and full transfer. Assisted transfer might connote situations where a person transfers their learning to the job, but does so with significant assistance, support, or prompting. A person who transfers their learning to the job because their manager strongly urges them to do so would achieve assisted transfer. A group of learners who transfer their learning to the job only because their course instructor is acting as an after-training coach would achieve assisted transfer.

Of course, most transfer is assisted in some way—with the learner given permission, time, or resources; however, the distinction between assisted transfer and full transfer is more subtle. Assisted transfer represents situations where the learner would have been unlikely to engage in transfer on their own without assistance, support, or nudging. Assisted transfer and full transfer achieve the same ends, but assisted transfer is more tenuous, more dependent on significant help from other people.

Full transfer occurs when a person takes what they've learned and successfully puts it into practice in their work without *the need* for significant help or prodding. Even if they got assistance, support, or nudging—if they would have put their learning into practice even without these interventions—we would designate the result as "full transfer."

Let me emphasize again that both assisted and full transfer represent complete success in transfer! Indeed, we might argue—when an individual is forced to work with others to achieve transfer—that *more* benefits are created, including such benefits as team building, social impetus to persevere, interpersonal bonding, and socially-

mediated learning. Full transfer gets a slight edge in status because it indicates the learner was fully prepared to transfer learning to work without additional intervention.[25]

Tier 8—Effects of Transfer

Transfer is often considered the penultimate goal of learning, because learning is almost always instrumental—a means to other ends. Even when we learn for the pure joy of learning, we can consider joy to be an outcome of learning. Given learning's instrumentality in creating other outcomes, our evaluation efforts should assess learning results.

For almost all workplace learning efforts, we invest in learning because we hope it will be instrumental in achieving other ends. We train nurses not just so they transfer their skills to their work, but because we expect to get better patient outcomes. We train managers in leadership skills not just so they perform better as managers, but because we hope their teams will achieve better results (for example, increasing product or service quality, reducing costs, and increasing employee morale and productivity). Even when we just provide training as an employment benefit—one that is intended to directly benefit our employees—we hope such organizational generosity might translate into increased loyalty, lower turnover, recruiting benefits, etc. The notion that learning should be instrumental is one of the great features of the Kirkpatrick-Katzell model. Level 4 (Results) is a clear statement that learning ought to achieve other ends.

Some have modified Kirkpatrick-Katzell's original Level 4 label "Results," changing it to "Business Results." Of course, business results are not the only outcomes we might hope for. Roger Kauffman, in particular, has emphasized the potential for assessing learning's impact on society. Elsewhere I've emphasized the importance of acknowledging the outcomes that affect the learner. In my mind, it seems we've had a blind spot, sometimes completely ignoring our learners as stakeholders in the learning process. This is especially apparent when we consider the psychological investment learners often make—first, challenging themselves to overcome their own weaknesses and limitations, then dealing with anxiety and resistance and blowback as they champion new practices in the workplace.

Transfer can affect many people and things. If we're conducting an evaluation, we ought to at least generate a list of the potential stakeholders and impacts. Here's the list of possible stakeholders presented in LTEM Version 13: beneficiaries, learners, the organization, customers, sponsors, coworkers/families/friends, investors, community, society, and the environs, etc.

[25] In LTEM Version 13, I have removed the distinction between full and assisted transfer from the one-page framework because I and others were not finding it critical to our evaluation work. Still, it is important to understand that transfer is not on or off, but that it is influenced by contextual supports and can require time and effort to be fully realized.

Learning transfer can produce benefits and harms. Too often, we focus only on the benefits of our learning interventions. We should also examine the downsides. We ought to look at time usage and productivity loss. We ought to look at the environmental damage from flying people across the country or the globe. We ought to look at the damage caused when we teach people nonsense. In the learning field, we harmed a whole generation of learning professionals (and their learners and organizations) when we taught learning-styles snake oil. Our exuberance in the early days of social media (for learning) led to bad information being shared online between our employees. We ought to look at the damage caused when we create compliance training for legal purposes that unintentionally sends messages that the organization is just covering its ass—that maybe having learners click "next" through an elearning course is all we need to do. We ought to look at the harm we might be doing to our fellow citizens when we teach union-busting tactics or teach managers how to squelch employee requests for better working conditions. We ought to examine the dangers of content that has no scientific backing or no vetting by true experts. We ought to calculate the waste that occurs when we throw training at non-training issues.

Enough! I made the point. We have a responsibility to think broadly about our learning impact and consider both benefits and harms.

Creating beneficial transfer effects usually requires more than just our efforts as learning professionals. Our learners must take responsibility for transfer. Their managers and coworkers must lend support or be cooperative. Organizational units might have to change practices and policies to enable success. Where training is part of a larger strategic initiative, many other factors will come into play.

To fully vet transfer effects, we must use rigorous methodologies to ensure we can separate out the learning effects from other factors that may influence the target results. This is not always easy or even possible.

Let's say an organization provides training to call center reps in attempts to (1) reduce the length of phone calls and (2) increase customer satisfaction with those calls. At the same time, the organization redesigns its product line to make it simpler and more user friendly. The result: The number of calls to the reps drops, as does the time needed in the calls—and customer satisfaction soars. Did training make a difference? Unfortunately, it will be impossible to tell which factors produced the results unless we had earlier decided to use a control-group design and withheld training from some of our reps. By comparing those reps who had gotten the training to those who hadn't, we'd be able to tell what effect the training had. This type of control-group design might be worth it—particularly in some strategically-important initiatives or to test our general training methodology—but, in most cases, organizations will not take the risk of withholding training from employees.

I'm oversimplifying here. Control group designs can provide training to all employees if we're willing to delay training for some employees. Further, while randomized control trials are the gold standard, there are other designs that can help us get a sense of whether it was the training that made a difference, including pre- to post-testing, time series analysis, comparing different learning designs, etc. When I teach about these options in my LTEM Boot Camp workshop, my participants see possibility but also worry about the complexity and investment involved. Measuring at LTEM Tier-7 and Tier-8 does involve substantial investments. Simple surveys do not cut it.

To summarize the key requirements of Tier-8 (Effects of Transfer), we should (1) consider learning outcomes that affect an array of stakeholders and their environs, (2) look for both benefits and harms, and (3) use rigorous methods to ensure we're able to draw conclusions about the causal impact of the learning experience itself.

The Benefits of LTEM

We've now looked in depth at each tier of LTEM—the Learning-Transfer Evaluation Model. It's now time to step back and think more broadly about LTEM in practice.

More and more organizations across the world are using LTEM and have found substantial benefits from it. Organizations report it:

- Nudges them to think beyond learner surveys.
- Enables them to build more rigor into their evaluation strategies.
- Brings a breadth of hope that learning evaluation can deliver results.
- Inspires their learning teams to substantially improve learning designs.
- Helps them discuss learning strategy at a higher level with their sponsors.
- Enables them to work more fruitfully with their vendors.

LTEM is designed to be practical. It does this first and foremost by sending clear messages and emitting nudges that can prompt you and your organization to action. So, for example, if your only evaluation metric is learner satisfaction (LTEM Tier-3B), you can see in the model that what you're doing is inadequate, so you should aim to do more. Many simple messages!

Note that LTEM offers five tiers of evaluation in which we can certify some degree of success in learning (Tiers 4A, 5, 6, 7, and 8).

- ***Tier 4A: Foundational Knowledge***
  Asks the question, *"Have learners fully comprehended the knowledge that is needed for real work decision making?"*

- ***Tier 5: Decision Making***
  Asks the question, *"Do learners know what to do when faced with relevant work scenarios?"*
- ***Tier 6: Task Performance***
  Asks the question, *"Can learners actually do what they learned to do?"*
- ***Tier 7: Transfer to Work Performance***
  Asks the question, *"In targeted work situations, are learners successful in using what they've learned?"*
- ***Tier 8: Transfer Effects***
  Asks the question, *"If learners have been successful in transfer, what effects has that transfer had on targeted and untargeted outcomes and stakeholders?"*

Note that LTEM has two tiers that are deemed completely inadequate in validating our learning results: Tier-1 (Attendance/Completion) and Tier-2 (Learner Activity). Let's be careful here. This doesn't mean we should avoid taking attendance or measuring learner activities! There are other reasons to pay attention to these metrics. For example, taking attendance can encourage learner engagement and give us an indirect indication of learner interest, motivation, and engagement—attributes we can measure to gain insights about our learning designs. Measuring learner activity during learning can also be useful in helping us diagnose strengths and weaknesses in our learning programs. So, while measuring at Tiers 1 and 2 may not enable us to validate the success of our learning, such measurement can give us useful insights for improving our learning designs. That is, they can help us with our formative evaluations.

Note that LTEM Tier-3 (Learner Perceptions) is divided into two approaches, one of which is deemed completely inadequate as a validation metric and the other of which is deemed generally inadequate. Querying learners about constructs that are unrelated to learning effectiveness is clearly inadequate—because learning effectiveness is our ultimate goal. On the other hand, querying learners about constructs related to learning effectiveness can give us some indication of whether the learning is likely to be successful, but learner feedback on its own is not sufficient to validate learning success.

LTEM is a roadmap. You don't have to take the whole journey in a day! Use LTEM to see where you are in terms of learning evaluation and see what you're missing. Make a plan to do better. We all work in situations that have constraints and goals that may limit our options. Your circumstances may not allow you to measure everything. Indeed, most of us (most of the time) will not be evaluating at all tiers in the model. We won't even be evaluating all the green-is-good validation-enabled tiers in the model. That's okay! Use LTEM to improve the feedback you get—to enable you and your organization to improve the learning you're creating.

Let me be even clearer. It's generally inadvisable to measure a learning program at all eight tiers of LTEM. There would be just too much investment involved. There are exceptions: for example, if you have a strategically important program or you want to measure at all eight tiers as a learning opportunity for your team.

If you're responsible for a large number of learning interventions, you don't need to evaluate each program every time you roll it out. This is too costly and, unless the learning designs are completely unique for each learning intervention, the insights you gather will be redundant.

When should you invest in deeper, more complete evaluations—that is, evaluations utilizing metrics from many tiers? With strategically important learning interventions, costly learning interventions, and learning interventions that will be deployed through many deployments or many years. You should use deeper learning evaluations when you are testing new learning designs—and periodically for all your important learning interventions.

You do not need to engage in deeper learning evaluations when you just recently evaluated a learning intervention using deep evaluation methods. If you basically use the same learning design in all your learning programs, you don't need to evaluate each one deeply. You might instead set up a rotating schedule of deep evaluations.

What if you're a vendor—for example, you create classroom or elearning programs for other organizations or individuals? The best vendors build evaluations into their learning solutions. They don't ask; they just incorporate deep evaluations into their projects. They do this to ensure they are providing maximum value to their customers. They create cycles of continuous improvement not only to benefit their customers but also to benefit their own organization. By enabling valid feedback, they enable themselves to create demonstrably superior learning programs that gain them customers.

Deeper evaluation can also be used to benefit your colleagues. When we work in a situation where feedback is missing (or worse, where it is invalid) we don't learn the right things. For example—and I've seen this firsthand—if we work for an organization that only looks at learner satisfaction data, we get faulty information and we learn to design and deliver learning that gets good satisfaction ratings but that almost certainly fails to be effective in supporting comprehension, remembering, and on-the-job application. By using deeper evaluation methods, we get more accurate information about the success of our learning designs. We begin to learn the right things: what really works and what doesn't. LTEM, then, can be used to educate you and your team on a continuous basis. When your team is smarter, they'll not only create better learning; they'll like their work better. When people like their work, they stay with their organization.

One of the dirty little secrets of the workplace learning field is that there is a wide disparity in the skillsets of those of us who are in it. Some of us are brilliantly effective because we use research-based wisdom and years of experience in looking at valid evaluation results. Some of us are dangerously ineffective because we follow recipes to guide our design work, tend to follow fads instead of evidence-based recommendations, and learn from faulty feedback derived from poor evaluation methods. If your organization is not using deeper evaluation techniques to gain validated feedback, you won't attract the best people, your team's conversations and interactions will focus on the wrong things, and you'll create a vicious cycle that reinforces poor learning-design thinking.

Worries about LTEM

I've been using LTEM, teaching about LTEM, and talking with learning professionals about LTEM since before its original publication in 2018—about eight years of conversations. I've heard it all. Most people resonate with LTEM and talk about its benefits. But some people say LTEM is too complex, too hard to communicate. Let's examine these concerns.

People have actually said to me that they don't think people in L&D are smart enough to use a model like LTEM. This really pisses me off because it's insulting to me and all my workplace learning compatriots. Also, it's simply not true that LTEM is too complex. Clearly, many people have already used LTEM to great benefit, so this worry is over the top. Also, LTEM fits on one page and has a very simple structure: eight simple tiers, with clearly stated labels. LTEM is not that complex. Indeed, compared to many other evaluation models that read more like process models, LTEM is easy.

Still, it is fair to be concerned about learning the new model, becoming as familiar with it as we are with the Four-Level Model, and being able to discuss it fluently with others.

It may take time for you and your team to get familiar and conversant with LTEM. For more than half a century, we've lived and breathed within the cozy confines of the Kirkpatrick-Katzell Four-Level Model. The Kirkpatrick-Katzell Model's great advantage—aside from its seminal notion that learning must produce results—is its simplicity. We can make sense of the four levels in a minute. The new model is more sophisticated, in keeping with the new professionalism needed in the rapidly advancing learning field. LTEM, because of its in-depth sophistication and its power to convey many more important messages, will require more from us. It will require us to study it until we comprehend it fully. It will also, when we use it, push us to create much better learning interventions.

What about conveying LTEM to others? Many have had success using the LTEM framework as is. Others have utilized my simplified diagram of LTEM—as available on the LTEM web page—or created their own simple depiction.

Recently, when someone worried how to communicate the essence of LTEM in a short elevator speech, I suggested something like this:

- "LTEM brings learning wisdom back into learning evaluation. By separating learning outcomes into knowledge, decision-making, and task performance (LTEM Tiers 4, 5, and 6), we raise the bar to more effective learning."
- "We don't settle for knowledge checks—and we all know knowledge is not enough."
- "Now we can focus on work-performance competencies like decision making and task performance."
- "LTEM not only improves what we measure; it elevates us to aspire to more effective learning."

Of course, I'm only emphasizing some of what makes LTEM unique and beneficial. Depending on your goals and your audience, you can craft your own elevator pitch.

Messages Shared by LTEM and the Four-Level Model

I recommend that LTEM replace the Kirkpatrick-Katzell Four-Level Model. But maybe you're still skeptical. Let me make the case another way. Let's examine which messages and nudges are shared by LTEM and the Four-Level Model. In the section after this one, I'll list the messages covered by LTEM but not the Four-Level Model.

Here are the messages and nudges they have in common:

- It is not sufficient to only measure LEARNER PERCEPTIONS.
- We should measure LEARNING OUTCOMES (knowledge, skills, abilities).
- Just because learners CAN PERFORM TASKS during a learning event doesn't mean they'll be ABLE TO CARRY THEM OUT in their work.
- Measuring WORK PERFORMANCE aligns with the central goal of learning: to enable learners to use what they've learned in their work.
- It is helpful to consider separately WORK PERFORMANCE and the EFFECTS OF WORK PERFORMANCE.

Messages Unique to LTEM

Here are the messages and nudges unique to LTEM—not specifically conveyed by the Four-Level Model:

- Just because learners ENGAGE IN LEARNING, or COMPLETE A LEARNING EVENT, or PAY ATTENTION, or SHOW INTEREST, or ACTIVELY PARTICIPATE doesn't mean they will have learned. Therefore, measuring these constructs is not adequate to validate our learning success.

- Just because learners say they LIKE A LEARNING EVENT doesn't mean they learned. Therefore, surveying learners on their general satisfaction—and on other factors not related to learning effectiveness—is an inadequate way of evaluating learning.

- Surveying learners on FACTORS RELATED TO LEARNING EFFECTIVENESS is better than surveying them on factors not related to learning effectiveness.

- Just because learners REPORT THEY HAVE EXPERIENCED EFFECTIVE LEARNING METHODS doesn't guarantee they learned. Therefore, surveying learners on their experience during learning, while it can provide some useful information related to learning design, must be augmented with objective measures of learning.

- Just because learners can RECITE FACTS AND TERMINOLOGY doesn't necessarily mean they know what to do.

- Some knowledge is TRIVIAL. It may be interesting (and even relevant), but, if it is not needed to help people make work-relevant decisions, we should not measure it to validate our learning success.

- Some knowledge is FOUNDATIONAL, and comprehending foundational concepts helps people make work-relevant decisions.

- Just because learners can DEMONSTRATE FOUNDATIONAL KNOWLEDGE during or right after learning doesn't mean they'll be able to REMEMBER that knowledge later.

- Measuring knowledge AFTER A DELAY is better than measuring it during learning or immediately after learning—when we are measuring for the purpose of validating our learning success.

- Just because learners can show they COMPREHEND FOUNDATIONAL CONCEPTS doesn't necessarily mean they can FULLY TRANSLATE their knowledge into work-relevant decisions.

- DECISION MAKING IS A BETTER METRIC than knowledge recitation or retention. Measuring decision making is also better than gauging learner perceptions.

- Just because learners can MAKE WORK-REALISTIC DECISIONS during a learning event doesn't mean they'll remember how to make those same decisions later. Therefore, measuring decision making during or soon after a learning event is an inadequate way of evaluating learning.

- Just because learners can MAKE WORK-REALISTIC DECISIONS doesn't mean they can perform successfully; they also must be fluent, appropriate, and timely in the way they utilize those decisions in action. Therefore, measuring for task performance (that is, measuring decision making in action) is better than measuring for decision making alone.

- Measuring WORK-REALISTIC TASK PERFORMANCE is generally a better metric than measuring decision making, because task performance is more authentic to work performance than decision making alone—specifically requiring a fuller accounting of contextual cues and demonstrations of fluency in action.

- Just because learners CAN PERFORM TASKS during a learning event doesn't mean they'll REMEMBER how to carry out these tasks later. Therefore, measuring task performance during or soon after a learning event is an inadequate way of evaluating learning. Ideally, we'd like to measure task performance some time after learning.

- In MEASURING WORK PERFORMANCE, we must utilize RESEARCH-LIKE METHODS to help us determine whether it was our learning intervention rather than other factors that enabled the work performance.

- It's important to consider all possible results of our learning interventions, WHETHER POSITIVE, NEUTRAL, OR NEGATIVE—avoiding bias, preconceived ideas, or dishonest manipulations.

- It is important to consider the FULL SET OF STAKEHOLDERS that may be impacted by our learning interventions. We don't have to measure the impact on each stakeholder group—not all will be central to our goals or responsibilities—but we ought to at least consider whether we should take various stakeholder groups into account. Here is a short list: beneficiaries, learners, the organization, coworkers/families/friends, customers, sponsors, investors, community, society, and the environs.

- There are benefits to USING SUBJECTIVE METRICS like surveys, interviews, and focus groups, but such metrics can also introduce biases, ignorance, faulty reasoning, and other cognitive deficiencies into our data. Subjective metrics must first be well designed. They must also focus respondents on constructs for which they have pertinent knowledge. Subjective metrics should be augmented with objective measures. Multiple sets of well-designed subjective metrics are better than one set of well-designed subjective metrics.

- There are a NUMBER OF GOALS WE SHOULD HAVE as learning designers, including supporting our learners in building comprehension, remembering, decision making competence, task competence, and perseverance in applying what they've learned to their job or other performance situations.
- It is NOT SUFFICIENT TO SUPPORT LEARNERS ONLY IN COMPREHENSION, which is what too often passes for learning. At a minimum, we also need to support learners in remembering and in decision making, and ideally in task performance as well.

I'm sure I've missed a few messages and nudges and have inadequately conveyed some of those that are included. The big picture is that The Learning-Transfer Evaluation Model sends many messages that are critically important in our work as learning professionals. These messages enable us to get better feedback about our learning evaluations and our learning designs. By using LTEM, you will gain greater clarity in thinking about learning and learning evaluation, your organization and learners will achieve greater results, and the learning industry will have a better way to benchmark learning evaluation strategies and learning designs.

Using LTEM for Learning Design and Development (Not Evaluation Per Se)

I'm a big believer in working backwards from ultimate goals to learning design decisions. Many others before me have advocated for this idea as well—most famously in recent times was Stephen R. Covey, who suggested that highly successful people "*Begin with the End in Mind.*" Yet, this wisdom may go back to time immemorial, as evidenced in the biblical quotation (Ecclesiasticus 7:36), "*Whatsoever thou takest in hand, remember the end, and thou shalt never do amiss.*" In learning evaluation, Donald Kirkpatrick said, "*Trainers must begin with desired results and then determine what behavior is needed to accomplish them. Then trainers must determine the attitudes, knowledge, and skills that are necessary to bring about the desired behavior.*"[26]

LTEM can be used as a guide to work backwards from the ultimate goals of learning. When it is used in this way, the process is not an evaluation process, but a project management process used specifically to develop learning programs. In using it in this way, LTEM makes clear the ultimate goals toward which we are working.

[26] Kirkpatrick, D. L. (1994). Evaluating training programs: The four levels. San Francisco: Berrett-Koehler. Page 26.

To use LTEM as a learning design and development tool, here is a process I recommend:

1. Start by looking at Tier 8. What ultimate results are you hoping to achieve? Consider all the possible results, not just "business results." Look at all the impacts you might have on a wide range of stakeholders. Look for both positive and negative effects. Look also for effects on the environs—on the situations, objects, and entities that might also be impacted by your learning programs.

2. Afterward, look at Tier 7. How will transfer be manifested? What transfer results are you hoping to achieve?

3. After you have a clear sense of your Tier 7 and Tier 8 goals, create a list of Evaluation Objectives—a clear statement of the various measures you will use to evaluate your program's success. For example, if you are training sales managers on how to coach their direct reports, you might specify a 5 percent improvement in sales numbers of the sales managers' direct reports as one evaluation objective (a Tier-8 Effect-of-Transfer objective). You might also specify an acceptable rating of the sales manager's coaching performance (from a survey of those they were assigned to coach) as another evaluation objective (a Tier-7 Transfer objective).

   In addition to these work outcome evaluation targets, you can also develop targeted measures for the other tiers in LTEM. Specify how you will evaluate Tier-6 (Task Performance), Tier-5 (Decision Making), Tier-4 (Knowledge, where meaningful), and Tier-3 (Learner Perceptions, being sure to focus on learning effectiveness). Where Tier-1 (Attendance/Completion) and Tier-2 (Learner Activity) metrics are useful to get feedback that can lead to learning-design improvements, you should specify those in advance as well.

4. After negotiating the acceptability of the aforementioned evaluation objectives with key stakeholders, begin the instructional-design process by determining the targeted performance situations and the targeted performance objectives within those situations. In other words, specify what the learners should be able to do and in what situations they should be able to do those things.

5. Once the performance situations and performance objectives are okayed by key stakeholders, you can begin designing and developing the learning program.

6. Preferably, you should use an iterative rapid-prototyping approach—quickly building and testing parts of your learning program AND utilizing Tier-1 to Tier-6 evaluation methods to formatively evaluate and improve your learning designs.

7. After sufficient iterations of development and testing, you can deploy your learning program and then evaluate its effectiveness using at least Tier-3A learner-feedback queries focused on effectiveness, Tier-4A (Foundational Knowledge), Tier-5 (Decision Making), Tier-6 (Task Performance), Tier-7 (Transfer to Work Performance), and Tier-8 (Transfer Effects).

8. Gather and analyze the data; share it with key stakeholders—especially the relevant instructional designers, managers, and trainers; determine plans for improvement; and then push forward with those plans while reinforcing lessons learned.

9. For strategically important learning programs, evaluate routinely. For other programs, evaluate periodically to ensure continuous improvement.

For brevity, the process outlined above is a foreshortened version of a fully specified project management process. Still, it incorporates several important principles. Start by focusing on the Work-Outcome tiers Tier-7 and Tier-8. From there, build a separate and distinct set of Evaluation Objectives—the metrics you will use to hold yourselves accountable. Consider all tiers of the LTEM model as you brainstorm evaluation options. Involve your stakeholders to negotiate agreements on your key objectives. Don't start instructional design work until you've clearly outlined your evaluation objectives and your targeted performance situations and objectives. Iterate and test, involving formative evaluation where appropriate. Evaluate continuously (or at least periodically) after your learning program is up and running.

Competitive Advantage from Learning Evaluation

In my latest book, *The CEO's Guide to Training, eLearning & Work: Empowering Learning for a Competitive Advantage,* the central argument is that we who work in learning and development can create for our organizations and our learners a competitive advantage. We do this when we use practices that are more effective, more innovative, and more powerful than those of our competitors.

Learning evaluation is a perfect target for competitive advantage. Most organizations are doing a seriously crappy job in evaluating learning. Most use poorly constructed smile sheets, which gives them bad data and nudges them to use learning designs that are inadequate or harmful. When our competitors are wallowing in the mud, all we need to do is take frequent showers to have a competitive advantage.

Squeezing more from this fanciful metaphor, if we want to maximize our advantage and create "barriers to entry" (as business strategists might say), we'd do more than take showers. We'd not only hire the best stylists to clean up our look, but we'd also educate ourselves to speak with eloquence. We'd practice and explore and deepen our

expertise. We would rise so far above the mud and the muck and the mire that our competitors would sink further into envy and remorse.

What I'm saying is that those of us who do a better job of evaluating learning—and use what we learn to create virtuous cycles of continuous improvement—are creating a competitive advantage. Those of us who push beyond easy improvements in learning evaluation create sustainable competitive advantages.

LTEM is clearly superior in this regard because it brings learning wisdom into our efforts to evaluate, design, and develop learning. It is not all we need, but it's the best place to start to move toward competitive advantage.

Afterword

I am grateful for the opportunity to engage in working on learning research and learning measurement issues. Without the help of my family, my clients, and those who have volunteered to collaborate on this work, it would be impossible. I'd also like to acknowledge all those who have laid the groundwork over many years for the work we do now.

Will Thalheimer
October 2024

Bibliography

Alliger, G. M., Tannenbaum, S. I., Bennett, W. Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50*, 341-358.

Anderson, L. W., Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*, 1st ed. Harlow, Essex, England: Longman Group.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. *Handbook I: Cognitive domain*. New York: David McKay Company.

Chargaff, E. (1986). Quoted in J. J. Zuckerman, 'The Coming Renaissance of Descriptive Chemistry,' *Journal of Chemical Education*, *63*, 830.

Jennings, B. K. (2007). On the Nature of Science, *Physics in Canada* 63, 7.

Kirkpatrick, D. L. (1959a). Techniques for evaluating training programs. *Journal of the American Society of Training Directors, 13*(11), 3-9.

Kirkpatrick, D. L. (1959b). Techniques for evaluating training programs: Part 2—Learning. *Journal of the American Society of Training Directors, 13*(12), 21-26.

Kirkpatrick, D. L. (1960a). Techniques for evaluating training programs: Part 3—Behavior. *Journal of the American Society of Training Directors, 14*(1), 13-18.

Kirkpatrick, D. L. (1960b). Techniques for evaluating training programs: Part 4—Results. *Journal of the American Society of Training Directors, 14(*2), 28-32.

Kirkpatrick, D. L. (1994). Evaluating training programs: The four levels. San Francisco: Berrett-Koehler.

Marzano, R. J. (2001). *Designing a new taxonomy of educational objectives*. Thousand Oaks, CA: Corwin Press.

Marzano, R. J., & Kendall, J. S. (2007). *The new taxonomy of educational objectives, 2nd ed*. Thousand Oaks, CA, US: Corwin Press.

Pollock, R. (2017). Personal Communication, November 2017.

Sitzmann, T., Brown, K. G., Casper, W. J., Ely, K., & Zimmerman, R. D. (2008). A review and meta-analysis of the nomological network of trainee reactions. *Journal of Applied Psychology, 93*(2), 280-295.

Thalheimer, W. (2016). *Performance-focused smile sheets: A radical rethinking of a dangerous art form*. Somerville, MA: Work-Learning Press.

Thalheimer, W. (2022). *Performance-Focused Learner Surveys: Using Distinctive Questioning to Get Actionable Data and Guide Learning Effectiveness*. Somerville, MA: Work-Learning Press.

Thalheimer, W. (2024). *The CEO's Guide to Training, eLearning & Work: Empowering Learning for a Competitive Advantage*. Somerville, MA: Work-Learning Press.

Uttl, B., White, C. A., Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation, 54*, 22-24.

The Learning-Transfer Evaluation Model

A Framework on the Outcomes of Learning. Version 13. *(Known as LTEM, pronounced "L-tem.")*

Will Thalheimer

WORK-LEARNING RESEARCH

Outcomes Measured In the Work

Tier		
8	**Effects of Transfer**	**Learners' work performance impacts KEY PERFORMANCE INDICATORS AND KEY STAKEHOLDERS.** Possible stakeholders: beneficiaries, learners, the organization, customers, sponsors, coworkers/families/friends, investors, community, society, and the environs, etc. *Tier-8 should utilize objective metrics and rigorous methods to determine the causal impact of learning—positive, negative, or neutral. Surveys almost always provide weak evidence for KPIs.*
7	**Transfer to Work Performance**	**Learners PERFORM WORK TASKS AND/OR CREATE WORK OUTPUTS using what they learned.** Validated by objective measures of task behaviors and/or outputs—or multiple subjective measures. *Tier-7 requires rigorous methods to determine the causal impact of the learning program—positive, negative, or neutral. KBIs (Key Behavioral Indicators) can be outlined as generally-worded goals but must be operationalized using valid constructs that are transparent and measurable.*

Outcomes Measured In the Learning

6	**Task Performance**	**Learners PERFORM RELEVANT REALISTIC TASKS.** Demonstrated without unrealistic hints or help. May involve multiple tasks chained together. • Assessed Shortly After Learning—measured during or right after learning event. *Not a fully adequate metric because learners may forget how to perform.* • Assessed for Remembering—measured after several days or more. *ADEQUATE TO CERTIFY SUCCESSFUL TASK PERFORMANCE.*
5	**Decision Making**	**Learners MAKE DECISIONS GIVEN RELEVANT REALISTIC SCENARIOS.** Scenarios should include both relevant and non-relevant information so that learners can demonstrate they know how to evaluate (make sense of) contextual work factors. • Assessed Shortly After Learning—measured during or right after learning event. *Not a fully adequate metric because learners may forget how to make these decisions.* • Assessed for Remembering—measured after several days or more. *ADEQUATE TO CERTIFY SUCCESSFUL DECISION MAKING.*
4	**Knowledge**	**4A. Learners answer questions about FOUNDATIONAL CONCEPTS—those critical to work decisions.** • Assessed Shortly After Learning—measured during or right after learning event. *Not a fully adequate metric because knowing concepts does not always enable performance.* • Assessed for Remembering—measured after several days or more. *ADEQUATE TO CERTIFY FOUNDATIONAL KNOWLEDGE.*
		4B. Learners answer questions on TRIVIAL KNOWLEDGE—info not central to work decisions. • Assessed Shortly After Learning—measured during or right after learning event. • Assessed for Remembering—measured after several days or more. *Inadequate measures because knowing facts/terminology does not usually enable performance.*
3	**Learner Perceptions**	**3A. Learners are queried in a way that REVEALS INSIGHTS RELATED TO LEARNING EFFECTIVENESS.** • Measuring: whether learners feel they know, motivation to apply, practice, after-learning support, etc. *Such measures can hint at outcomes but should be augmented with objective metrics.*
		3B. Learners are queried about CONSTRUCTS NOT RELATED TO LEARNING EFFECTIVENESS. • Measuring: learner satisfaction, course reputation; or using poorly-designed metrics. *Such measures are inadequate to validate learning success.*
2	**Learner Activity**	**Learners ENGAGE IN ACTIVITIES RELATED TO LEARNING.** Measures may focus on learner attention, interest, participation, or other learner actions. *These measures are not adequate to validate learning success because learners may engage in these activities but not learn, or learn wrong information, or be unmotivated to apply what they've learned.*
1	**Attendance/ Completion**	**Learners ENROLL, START, ATTEND, OR COMPLETE A LEARNING EXPERIENCE.** *These measures are not adequate to validate learning success because learners may engage in a learning experience but not learn, or learn wrong information, or be unmotivated to apply what they've learned.*

 See next page for recommendations.

The Learning-Transfer Evaluation Model

Helpful Hints on How to Use LTEM

1. **LTEM** is a conceptual framework enabling flexibility in how things are measured. You will need to select valid metrics for each tier within which you plan to measure.

2. **LTEM** is designed to measure factors in learning transfer. It has learning wisdom baked into it. For example, it recognizes the importance of measuring remembering after a delay. It suggests we measure more than knowledge, specifically decision-making and task performance.

3. **LTEM** discourages the use of measuring attendance, completion rates, learner activities, and poorly-designed smile sheets as a way to VALIDATE our learning results. These are deemed inadequate to validate learning, because learners may seem successful on these measures but (a) may not have learned, or (b) learned wrong information, or (c) be unmotivated to apply what they've learned, or (d) be unprepared to take action, etc. We can use these metrics for FORMATIVE purposes—that is, to assess for ourselves what is happening in the learning.

4. **LTEM's** color scheme has meaning. Red table row measures are NOT AT ALL ADEQUATE to certify or validate learning success (Tiers 1, 2, 3B). Yellow row measures are NOT FULLY ADEQUATE to certify success (Tiers 3A, 4B). Green row measures are ADEQUATE TO CERTIFY SOME LEVEL OF LEARNING SUCCESS (Tiers 4A & up), though no one metric tells the whole story.

5. **LTEM** conveys the complexity of learning evaluation. Evaluation can be straightforward, as when we survey learners. But to more fully understand learning outcomes, we must engage in a process of discovery, one requiring rigorous methods, expertise, and significant time investment.

6. **LTEM does NOT** suggest that we measure at all eight tiers. Learning measurement takes time and resources and should only be undertaken where there are benefits. We must prioritize.

7. **LTEM does NOT** suggest we start at the first tier and move upward. Indeed, it specifically discourages us from trying to validate our learning at Tiers 1, 2, and 3B.

8. **LTEM does NOT** suggest we always aim for measuring at Tiers 7 and 8. Measuring work performance or the effects of transfer requires rigorous practices involving substantial time and investment. Measuring at Tier 7 or Tier 8 may simply not be worth the investment and can be counterproductive if we fail to measure learning at Tiers 4, 5, and 6—because foundational knowledge, decision making, and task performance provide higher-fidelity information about whether our learning programs were well designed.

9. **LTEM is NOT** a survey methodology. Will Thalheimer, LTEM's creator, has also developed the Performance-Focused Learner Survey approach; but they are separate things.

10. **LTEM has many uses, not just for learning evaluation.** It can be used for (a) learning-evaluation gap analysis (examining how you're evaluating now and what you want to do in the future), (b) planning your overall evaluation strategy, (c) credentialling courses, (d) motivating improved learning designs, (e) having sales discussions, (f) brainstorming and prioritizing evaluation options, (g) outlining evaluation goals at the beginning of development, (h) negotiating/discussing possible program outcomes with sponsors or other key stakeholders, and (i) framing evaluation methods to maintain clarity within a complex evaluation project, etc.