

New Paradigms Call for New Evaluation Methods: Moving Beyond Kirkpatrick

Lara Bove, Anne Little, Ph.D.
SAIC
Reston, VA
lara.k.bove@saic.com, anne.m.little@saic.com

ABSTRACT

While instructional systems design is the accepted standard for developing and delivering training, many organizations fall short with respect to the evaluation component (Keen & Berge, 2014). In truth most organizations follow ADDI, not ADDIE. This lack of appropriate evaluation or complete omission poses a critical problem because evaluations are supposed to help organizations determine the impact of training on job performance (Davis, 2016; Griffin, 2011).

The impact of training is not the most significant question. We want to know: Are people performing as expected? If not, why not and what can we do to change that? As technology improves so, too, does our ability to collect the performance data that will enable us to answer those questions.

Many training evaluation methods, including Kirkpatrick, are insufficient because they were not created with consideration to the types of data we now have available. Further, there are many misconceptions about the relationship between training and workplace performance (e.g., the commonly held myth that completing a task in learning automatically translates to successful performance on the job) which have led to poor evaluation practices. This paper will provide guidance for selecting and designing evaluation methods that are appropriate for the given training context as well as methods for identifying and collecting the appropriate data. The paper will also look at specific types of workplace behaviors and identify theories which can support the evaluation of training.

The field of training evaluation is still in the development stages, and has not come to a consensus on the methods for evaluating learning within organizations or the aspects of the training that should be evaluated. Identifying evaluation methods which utilize the data available can enable organizations to make more informed decisions about the needs of our workforce (whether training or otherwise).

ABOUT THE AUTHORS

Lara Bove is a Senior Instructional Systems Designer with SAIC. She has more than twenty years of training development experience, and her research interests include evaluation of training and adaptive training. Ms. Bove has designed training for federal clients, state governments, and the Department of Defense, and include the Federal Aviation Administration and the Department of Veterans Affairs. Ms. Bove holds an English degree from the University of North Carolina at Greensboro and an MS Ed from James Madison University.

Anne Little, PhD is Vice President of Training Solutions Development with SAIC. Dr. Little has more than twenty years of training development and delivery experience, and her research interests include motivational strategies within online learning environments. She has leveraged her experience to design training programs for numerous federal clients, including the Department of Defense, Department of Homeland Security, Department of Energy, and the Federal Aviation Administration. Most recently, her research and development portfolio has focused on re-engineering compliance training programs. Dr. Little holds a Mathematics degree from Purdue University, and a PhD in Instructional Technology from George Mason University.

New Paradigms Call for New Evaluation Methods: Moving Beyond Kirkpatrick

Lara Bove, Anne Little, Ph.D.
SAIC
Reston, VA
lara.k.bove@saic.com, anne.m.little@saic.com

OVERVIEW OF THE PROBLEM

The U.S. military is implementing several transformative programs to modernize the force, including the Navy's Sailor 2025 initiative and the establishment of Army Futures Command. The focus of Sailor 2025 is on "empowering Sailors, updating policies, procedures, and operating systems, and *providing the right training at the right time in the right way to ensure Sailors are ready for the Fleet*" (emphasis added). Army Futures Command describes itself thusly, "Above all else, we want to make sure Soldiers have what they need, before they need it, to defend tomorrow...today." The Air Force's AFWERX seeks to remove acquisition road blocks and support rapid acquisition for research-based development, thereby "improving Air Force capabilities by connecting innovators and accelerating results." Clearly training will play a critical role in these transformation efforts.

Ultimately, all of these efforts demonstrate an understanding of the need for a paradigm shift in the way we support our forces. And these transformation efforts indicate a willingness to consider the validity and effectiveness of the processes and tools that we are currently using. We submit that the evaluation tools which are currently used for training are ineffective and can actually damage these transformative efforts.

In order to provide training support for the transformation of the force as well as the non-military workforce, we must use more than the latest technology or the most creative ideas, we must also apply the latest in scientific approaches and understanding of human behavior, learning theories, data analytics, and evaluations. We must do more than provide our employees with the training they need. We must also evaluate our training to ensure that our workforce is using what they've learned on job (or on the battlefield, in the skies, and at sea). Unfortunately, many organizations are not aware of the fact that the evaluation methods they are using do not work, which means they do not have an accurate picture of the effectiveness of their training. This paper recommends replacing Kirkpatrick with a research-based evaluation model and provides a structured approach to implementing the new evaluation model.

Research-Based Practices in Training

In terms of instructional approaches, the Instructional Systems Design (ISD) model is research-based and has been proven effective since its development in 1970s. When we use research-based methods we can predict the outcomes of our training with relative confidence. The ISD model also enables us to figure out what went wrong when things do not work as we expected: if we follow the model, including evaluation, and workplace performance does not change, then we can have some confidence that the problem lies outside of the training. Perhaps it was not a training problem after all. Or perhaps there were barriers to training transfer which we did not account for.

The ISD model is comprised of five phases, most commonly identified as Analysis, Design, Development, Implementation, and Evaluation (Hannum, 2005). Figure 1 depicts one common visualization of the ADDIE model. While the process generally begins with an analysis and ends with evaluation, the five phases are iterative and all

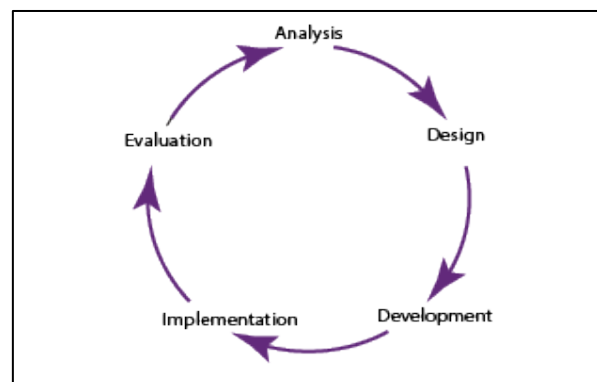


Figure 1: ADDIE Model

phases are critical to the integrity and effectiveness of the training. The key point is that “effectiveness” is related to how well training enabled learners to meet the objectives. And this is where the disconnect begins because the ISD process focuses on meeting the learning objectives (as opposed to workplace performance). Despite the widespread acceptance of the ISD model, many organizations are leaving evaluation out of the process (Keen & Berge, 2014). Indeed, the federal government’s own Advanced Distributed Learning Initiative cites a lack of evaluation of the impact of training as one of the top 24 problems in Department of Defense (DoD) training, and notes that government contractors are not evaluating the training which they develop (Advanced Distributed Learning Initiative, 2017).

Still others fall into a trap of conducting quasi-evaluations which is to say they “evaluate” the training, but they are following methods which provide inaccurate information about the training’s effectiveness, do not include a plan for addressing any shortfalls in the training, and do not provide insights about training transfer.

We might say that such organizations are not following the ISD model, but rather they are following an ADDI model, which is neither research-based nor a proven methodology. The problem is that without evaluations we cannot know if our training is really working. A quasi-evaluation is even more dangerous because people put stock in something that has no basis in reality.

In addition, without a clear link between the training and its outcomes, an organization can devalue the work of its training department and view training as a cost with little or no value (Griffin, 2011). If the training’s value is not recognized, then the organization may not spend the needed resources to ensure the training is successful (Barnett & Mattox, 2010). In a sense, the problem of not evaluating the training leads to a downward spiral of less and less effective training (even with good quality), and poorer outcomes on the battlefield. By not evaluating the outcomes to determine the training’s effectiveness, we cannot know if the training worked. Further, if we don’t know if it worked, we may not think the training had any value at all. And, if the training has little value, then we don’t invest much to ensure effective training. Further still, if we don’t invest in effective training, we won’t produce effective training. Our workforce, and our warfighters, do not get the training they need, and they are less prepared. And the downward spiral continues.

CURRENT TRAINING EVALUATION METHODS

In order to determine best practices for evaluation and unravel the disconnect in practice, we begin by clearly stating the purpose of training evaluations. An *evaluation* is a systematic process for collecting data related to questions about a program with the express purpose of increasing our understanding and using the information to inform decisions about the program (Russ-Eft & Preskill, 2009). Thus, *training evaluations* must meet all of these criteria:

1. follow a systematic process
2. seek to answer specific questions about the training’s effectiveness
3. have the express purpose and intent to take action based on the findings

Anecdotal evidence has led the authors to believe that many ISDs are unaware of the requirements articulated by the last requirement.

Unfortunately, there are strongly entrenched evaluation practices in place that have led organizations to believe that they are getting good evaluative data because their instructional systems designers or training specialists are following methods which are inherently flawed and provide inaccurate information about the training’s effectiveness. This problem was bound to happen because the most common method of evaluation in place today (Kirkpatrick) was implemented before the research-based ISD model was developed. In addition, it was difficult to collect the kinds of data that we can now use to measure the impacts of training on the workplace. During the early 20th century the Department of Defense and U.S. government focused efforts on developing methods for training adults (Knowles, 1980). Concurrent to this, evaluations were becoming more and more important to the military and federal government (Russ-Eft & Preskill, 2009). As an example, during World War II the military sought methods to test the abilities and aptitudes of new recruits, and looked for better practices to train them (Russ-Eft & Preskill, 2009; Knowles, 1980). Even as the field of adult training was continuing to develop, the demand for programmatic evaluations grew. So, in 1959, Kirkpatrick first published his ideas about training evaluation (Kirkpatrick, 1996); however, the ISD model would not be developed or introduced to the public until almost 20 years later (Hannum, 2005). At some point, the ISD model was combined with Kirkpatrick. After all, the ISD model called for evaluation and Kirkpatrick provided

one. Unfortunately, because ISD is research-based some training practitioners have assumed that Kirkpatrick is also research-based and provides the same level of assurances as ISD. This is not correct.

The authors have also found current evaluation methods to be limited in their ability to connect training to its impact on the job. In response to this problem, a research effort examined evaluation methods which would link training with workplace outcomes as well as methods for identifying and collecting the data which would support such evaluations. The meta-analysis of evaluation methods for compliance training found that there is little in the way of structured evaluation methods for practitioners (Bove, 2018).

Kirkpatrick's Four Levels and Quasi-Evaluation

As previously mentioned, Kirkpatrick's Four Levels of Evaluation were initially identified in 1959: Level One-Reaction, Level Two-Learning, Level Three-Behavior, and Level Four-Results (Kirkpatrick, 1996). Kirkpatrick's Four Levels are now a common and accepted training evaluation method (Barnett & Mattox, 2010; Holton, 1996). However, researchers have expressed several concerns and identified many limitations with Kirkpatrick's Four Levels (Griffin, 2010; Holton, 1996; Kaufman & Keller, 1994).

One significant criticism is that the model has been accepted as if it were based in research and theory, but in fact it is a taxonomy, or a way organizing different types of evaluation (Holton, 1996). The difference is critical because there is no research to support the notion of a hierarchical ordering among the four levels. Nor has research demonstrated causality among the levels. Yet, in practice, people use Kirkpatrick's Four Levels and make decisions about training and organizations as if such relationships did exist (Holton, 1996).

For example, years of research have not shown any connection between a user's satisfaction with the teaching method (Level One) and their actual mastery of the content (Level Two) (Hannum, 2009b). In addition, Level One evaluations do not help us determine the direct impact of training on the organization (Alvarez, Salas, & Garofano, 2004; Haskins & Shaffer, 2011). Still, among organizations which conduct evaluations, most are limiting their evaluation to a Level One (Wang & Wilcox, 2006). What if these organizations are making decisions about the value of the training (or perhaps even the effectiveness of the training) based on user satisfaction? We must have the courage to say the goal of the training is not to make a learner happy; the goal of training is to ensure workers are prepared and able to do their job in accordance with the organization's stated values and mission. In the military context, the goal of training is to ensure our warfighters meet the mission and come back home. Consider BUD/S (Basic Underwater Demolition/SEAL) training; it is not likely that the training methods in place today would meet the criteria for user satisfaction.

It is true that Level Three evaluations focus on whether the training is impacting job performance and Level Four evaluations look at how the training has impacted the organizations. However, in practice, among organizations that conduct evaluations, only 12 percent evaluate the impact of the training on job performance (Wang & Wilcox, 2006; Griffin, 2011). Further, in terms of evaluating training and looking at impact on the organization, Kirkpatrick's Four Levels do not provide enough guidance to address the many factors which influence whether someone uses what they learned in training when they are on the job. This deficit is not surprising; models that follow research-based paradigms would seek identify and separate confounding variables. In fact, Holton recognized this problem and began conducting research to separate, identify, and validate the confounding variables which impact whether training is transferred to the workplace (Holton, 1996). Holton's work led to the development of the learning transfer system inventory (LTSI), which is discussed in the next section.

Without causality or a hierarchy among the Four Levels we must recognize that Kirkpatrick's Four Levels do not meet the evaluation criteria of following a systematic process. Further, while organizations may use Kirkpatrick's Four Levels to answer questions about the training's effectiveness, the information generated from a Kirkpatrick-based evaluation does not enable such answers. Thus, the second criteria for training evaluation is not met. As for the third criteria, Kirkpatrick's Four Levels does not specify that the organization include plans to do anything with the information collected. As Russ-Eft and Preskill (2009, p. 24) so elegantly stated, "If there is no intention to use the findings, then an evaluation should not be undertaken."

Learning Transfer System Inventory (LTSI)

Since the goal of training is to improve individual and organizational performance, evaluations should analogously measure performance on the job at the individual and organization levels. Measuring job performance is different from measuring whether or not people learned something during training; but this difference is frequently collapsed. We must measure performance in the workplace because there is no guarantee that people will incorporate what they learned into their work. Unfortunately, some training paradigms have wrongly equated training performance with workplace performance. Of course, if trainees don't learn it, then they can't apply it. But learning is not enough. The use of knowledge, skills, and/or abilities learned in training events on the job is known as "training transfer" (Blume, Ford, Baldwin, & Huang, 2010). If organizations want to measure training transfer, they must use a method which measures what is happening on the job as well as the factors impacting training transfer. Consider an example of training forklift drivers who operate in a warehouse. The drivers receive training to learn forklift procedures and appropriate route of travel in the warehouse. The training evaluation includes using GPS data which tracks driver routes in the warehouse (i.e., the workplace). Suppose the GPS data revealed that drivers were not following the routes they were taught in the training. Upon further investigation, it appears as though drivers knew the correct route, but were taking the shortest (most direct) routes instead. The drivers may have believed they were improving efficiencies; however, the purpose of establishing the routes and prescribing compliance exactly as taught was due to safety regulations. This example illustrates the disconnect from "learning from the training event" and "actually doing the correctly trained behavior on the job." Most organizations who fail to examine the performance data collected on the job, would likely be tempted to require the workforce to repeat the training because they do not consider other possible sources for performance deficits. By ignoring the organizational factors which impeded that training transfer, the organization's performance problems will persist. Instead of sending drivers back to repeat the training, the organization could change the environment, such as by putting up one-way signs or other direction indicators. This would solve the performance problem by a combination of well-designed training and implemented organizational change identified by the performance based evaluation.

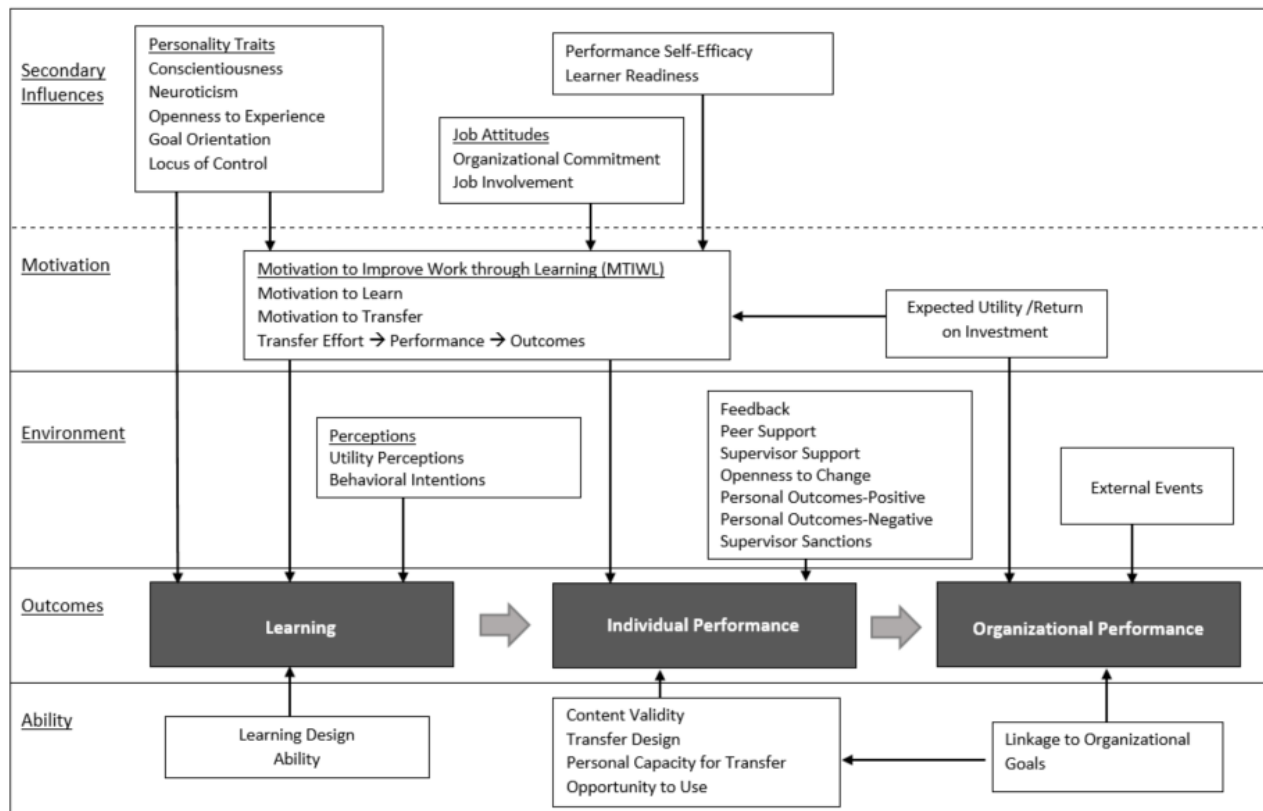


Figure 2. Holton Evaluation and Research Model (Holton, 2005)

As noted, Holton worked to identify the variables which impacted training transfer. Holton's research led to the development of the Learning Transfer System Inventory (LTSI), which is a research-based method for evaluating the impact of training through the lens of training transfer. The LTSI measures three separate outcomes: learning, individual performance and organization performance. Returning to the forklift training example, the learning outcomes are demonstrated mastery in the training, the individual outcomes are the driver's routes, and organizational outcomes are the routes that all drivers took, and how those impacted the organization. There are likely broader questions related to production such as individual and organizational metrics (e.g., how many pickups or deliveries per hour, number of accidents (and subsequent costs related to injury, equipment repair, or loss of product). It is important to distinguish the performance of the trained skills or practices so that the organization can determine whether those skills are being implemented.

Figure 2 presents the latest version of the Holton Evaluation and Research Model, which depicts the components of the LTSI. The diagram shows factors which have a direct effect on training transfer: motivation, environment, and ability. The secondary influences are factors which impact the primary factors, and can be thought of as moderators and mediators. For example, job attitudes impact an individual's motivation to learn or motivation to transfer, and that motivation impacts whether the individual uses the training on the job.

Through years of research and refinement, Holton and Bates developed the learning transfer system inventory (LTSI), which "is defined as all factors in the person, training, and organization that influence transfer of learning to job performance" (Holton, 2005). The LTSI, now in its third iteration, has been tested throughout the world in various workplace and training settings, and is grounded in theory and research (Bates, Holton, & Hatala, 2012). The LTSI is comprised of surveys that are sent to learners and their managers. The surveys are designed to measure a specific set of 16 factors, which comprise the different primary and secondary influencers that have been empirically shown to be interconnected and affect the outcomes of learning, individual performance, and organizational performance (Bates et al, 2012). The surveys are self-report and use a 5-point Likert-type scale (strongly agree to strongly disagree).

Since the survey is copyrighted, we cannot provide actual survey questions in this paper. However, here are the conceptual ideas that would be covered in questions on a survey sent to the employee, specifically related to Environment.

- The utility of the training: Does the employee think the training will help them on the job? How much of the training will be helpful on the job? How often is the learner faced with situations that will require them to use what they learned?
- Behavioral intentions: Does the employee intend to use what they learned when they go back to work? Are there parts of the training that they would never implement?
- Supervisor support: Does the employee feel that the organization or supervisor is in agreement with the approaches that were taught in the training?
- Peer support: Do your coworkers support you changing the way you do things? Are your coworkers open to change? Are there negative outcomes if you use the training?

The LTSI and Holton Evaluation and Research Model show that limiting an evaluation to student performance in the training environment does not provide enough information to determine whether the training will have the desired impact on job performance. If we want to figure out whether our training is impacting workplace performance, we must change our approach.

WORKPLACE PERFORMANCE VS. TRAINING PERFORMANCE

We acknowledge that there are times when training is not meant to solve a performance problem (in which case there would be no need to include a performance-based evaluation method). For example, training might be provided to help employees improve their skills using common software such as Microsoft Word and help people to function more efficiently, even if there is not a specific organizational problem with the way people are currently using the tool. However, much of the training we develop is designed to address an organizational problem that stems from lack of knowledge, skill, or ability. (E.g., the forklift drivers needed to know the routes as well as develop skills and abilities to drive the forklift). When we design training to enhance learner knowledge, skill, or ability, the training must focus on performance which demonstrates the knowledge, skill, or ability. In order to illustrate this, we present conceptual diagrams showing the difference between training-focused instruction and performance-based training.

Figure 3 presents two different diagrams of instructional design (ID) with a focus on learning outcomes, and not workplace performance. (Recall that learning outcomes are one of the three outcomes identified in the LTSI). 3A shows what happens in actual practice, where Evaluation is left out altogether; and 3B shows what practitioners believe they are doing. As an example, consider training which teaches a pilot how to communicate with air traffic control. There is a set vocabulary and phrasing which pilots must learn. There are specific procedures that must be followed. And there are motor skills and steps, such as ensuring the button is depressed when speaking. We could develop practice in the classroom, and even assess learners based on their responses in simulated environments with peers and instructors. We could include written tests for vocabulary or phrasing. The evaluation of the training would focus on the analysis, design, development, and implementation in the training environment alone. Of course, we need to be sure that pilots have mastered all of the materials. However, this training approach does not provide any insights into whether these pilots integrate everything that they learned when they are in flight, and whether there are any barriers to transfer (e.g., problems with equipment, cultural barriers, negative sanctions, and so on). There are those who might believe that pilot communication does not require workplace performance evaluations because there are other protections already in place. The thinking goes something like this: “Air traffic controllers and pilots must communicate in a certain way, and violating the standard operating procedures could result in difficulty communicating while in flight and the imposition of severe penalties from Command (or in commercial flight, from the Federal Aviation Administration (FAA)).” However, there are always factors which can impede training transfer. No performance components are immune to the risks inherent in organizations with human beings--not even standard operating procedures or so-called basic performance/maneuvers. We need only look to recent U.S. military mishaps such as the accidents in the Navy’s Seventh Fleet to recognize the severity of consequences of ignoring such risks. In 2017, there were two separate collisions involving Navy destroyers. “In both incidents, sailors on the bridge failed to sound a ship-wide alarm notifying the crew of danger, which is a standard Navy procedure” (Larter, 2017). The causes of the collisions notwithstanding, the fact that the sailors did not follow the standard procedure of sounding an alarm likely cost people their lives and resulted in more injuries. We raise the Seventh Fleet example with extreme sensitivity, and we are not suggesting that training evaluations are a tool which can help identify all of the problems within an organization. We do want to state that we must be sure we are not incorrectly assuming that success in training is equal to and guarantees successful job performance. Getting this right is a serious matter: we must ensure our workforce and warfighters are truly performing as trained.

As we think about performance, it is critical to distinguish mastery in training from job performance. It is not enough to apply the new skill or knowledge in the classroom. The LTSI demonstrates that there are many factors which influence whether mastery in the training environment is carried over and fully implemented in the workplace. But what about when the training environment closely lines up to, or even mirrors, the work environment, as is the case with simulations? After all, training needs to sync up with job performance (Hannum, 2005). And simulations enable us to provide training that closely resembles, if not mirrors, the work environment. We might even say that simulations are the best-in-breed of training. However, the training environment is not the same as the workplace. The people in the training context are not necessarily the same people encountered on the job. The instructor is not the same as the managers or commanders. The culture in the classroom may not be the same as what the worker faces on the job.

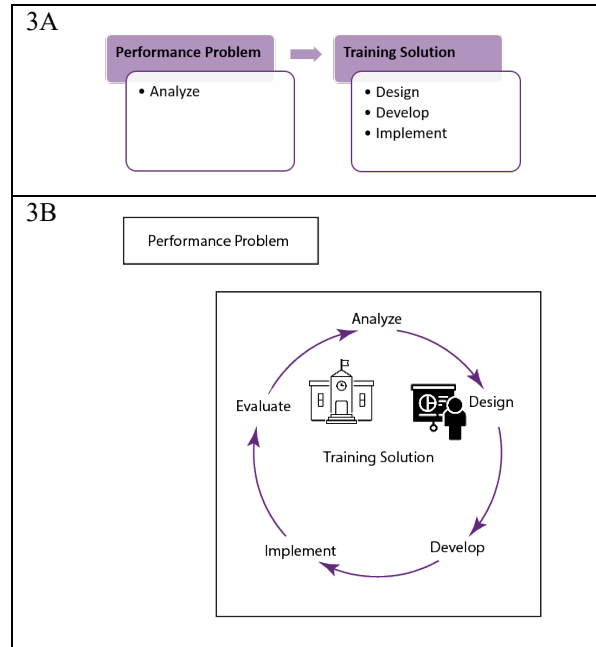


Figure 3. ID Focused on Learning Outcomes

3A: Analysis of a performance problem leads to a training solution. Training practitioners design, develop, and implement the solution. The process is linear: and even if evaluation is included, training is not improved based on findings. **3B:** A performance problem is found to require a training solution. The training practitioner analyzes, designs, develops, implements, and evaluates the training solution. The Evaluate phase informs future versions of the training solution.

A flight simulator allows the training pilot to practice dangerous moves in a safe environment while building requisite skills, reflexes, and thinking patterns. Virtual-live gaming situations allow our warfighters to practice and develop warfighting strategies and tactics. One might argue that this alignment between training performance and job performance means that training success is equivalent to, and guarantees, success on the job (in the air, at sea, on the battlefield). This is a grave mistake. There are three main limitations with this assumption:

1. It is possible that there are gaps in the training or areas where the training does not sync 100% to the live environment. Evaluating job performance can help us to identify those gaps so that we can remedy the training and better prepare the workforce and warfighter.
2. It is possible that environmental factors such as negative peer or manager support will prevent learners from using what they learned on the job.
3. It is also possible that some learners will misuse what they learned. Learners may apply principles in error or make mistakes we could not (or did not) account for in the simulation. We must determine that our successful learners are indeed applying the learning to the workplace.

Incorporating evaluation of job performance can help us to ensure that our workforce is truly prepared for the challenges they face on the job, and that they are using their training to its fullest advantage.

When the training only evaluates success in the classroom (i.e., training setting), we do not know if we have solved the organizational problem. While it is important to ensure that employees learned what was taught, the evaluation needs to look at performance on the job. It's not about training, it's about performance. Using the language of the LTSI, we must move beyond a focus on the learning outcomes to focus on the performance outcomes (individual and organization). Figure 4 depicts instructional design which is focused on workplace performance outcomes (the individual performance and organization performance outcomes noted in the LTSI). Returning to the example of the air traffic control communications, the training would address all of the knowledge, skills, and abilities required for the job as well as providing supports to ensure the highest levels of training transfer. This would include an evaluation of the training through the lens of training transfer which identifies barriers to implementing the skills.

The training would then put structures in place to address those barriers and ensure training transfer.

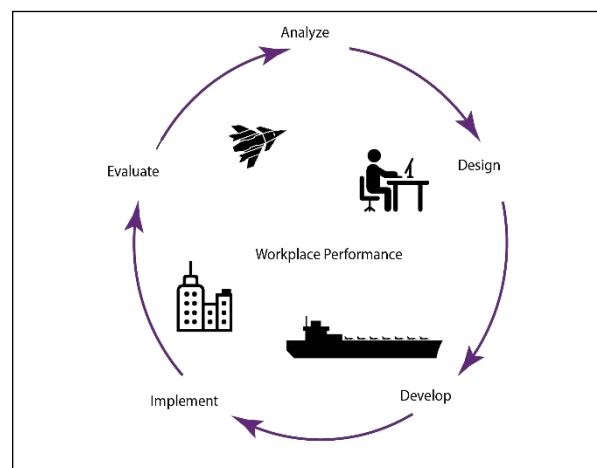


Figure 4. Workplace Performance Instructional Design

Here is an example of an evaluation that helped a telecommunications company identify barriers to transfer for a specific compliance training issue (Rustici, 2014). The company discovered that employees knew the correct, or expected answer, to the assessment question that was part of their training. However, given an opportunity to describe what they actually would do in the situation, a large number of trainees indicated they would not follow the company policy. The employees felt very strongly about the disconnect in terms of what they were expected to do, and what they thought was right. In short, it became clear that the employees had the best intentions of protecting their customers, and actually believed that violating the policy (and, in this case, the law) was within the customer's best interest. The evaluation of the training provided the company with insights they could use in addressing the issue—they had identified an important and specific barrier to training transfer. The company discovered employee intentions in the training context and was able to address the problem before employees took illegal actions. Without the evaluation, the company would have believed that employees were properly trained even though several employees intended to go against the training. Those same employees may have gone on to break the law, possibly having their actions go unnoticed for a great length of time until, and unless, there was an audit or some legal action was taken against the company.

THE EVALUATION METHOD

Since the purpose of the evaluation is to make decisions about the training, we must use methods that lend credibility to the results of the evaluation and allow us to draw conclusions about causality and correlation which we can use to

make decisions about the workforce. The evaluation method must also clearly explain how the training will lead to the desired performance outcomes, and explain how to measure whether the training met the goal (Rosi, et al, 2004).

Follow a Systematic Process: Incorporate the LTSI

In keeping with the three criteria for training evaluations which we identified earlier, we recommend incorporating the LTSI as the systematic process. The evaluation plan must incorporate LTSI surveys. Determine who will be surveyed, when the surveys will be sent, how you will collect and analyze the data, and who will be responsible for developing an action plan in response to the discoveries from the surveys. In addition to using LTSI surveys, the organization should consider the different influencers and design a training implementation plan which seeks to address barriers and enhance positive influencers. As with all of the process, the organization should have a plan in place to evaluate how well it worked. Identify metrics and collect data.

- What are some barriers to using the training on the job? (Will employee take longer to do their job at first? Do you expect to see increased hours with a decline as skill improves)
- What types of peer support or supervisor support will be needed? (How can the organization build this in and measure it?)

Seek to Answer Specific Questions About the Training's Effectiveness

The evaluation must seek to answer specific questions about the training's effectiveness. If we look back at the LTSI, there are three main outcomes from training: *learning, individual performance, and organization performance*. Ensure that the evaluation includes measures for each of these outcomes.

In order to ensure that the training and the evaluation are focused on performance, the training must be designed to address several overarching questions:

- What is the desired workplace performance?
- How can we prepare people to perform on the job?
- How can we measure performance on the job? What are the metrics we should use?
- What does "proper performance" look like? How will we know when people are doing what we expect?
- What are some of the barriers to proper performance?

Here are additional questions which can help to define what good and right look like in in the workplace. Notice that including metrics for organization performance outcomes expands the view of the impact of training.

- What will change if the pilot returns to the force and incorporates what they learned in training? (How is their performance going to be different as a result of the training?) What are the individual performance metrics we can collect from job performance?
- What metrics do you expect to remain the same before and after the training?
- What will change if most (or all) of the pilots incorporate what they learned? How will the organization performance change as a result of the training? What are the metrics we can collect for the organization?

Let's think about how this applies to developing training for flight simulator scenarios. We follow a progression which begins by determining the skills, behaviors, thinking patterns, and problem solving strategies our pilots need. We determine what right and good look like, using clear metrics (e.g., the pilot will do a CLES check (checking in cockpit to ensure all loose items are stowed) before performing aerobatic maneuvers in which the plane will be upside down). We develop a plan to teach those skills so that the training performance mirrors workplace performance as closely as we can. In this case, we use a simulator and build scenarios which put the pilot in different situations where s/he must remember to do the CLES check, and perform it properly. As we build the training and we determine training metrics, we should be thinking about metrics for workplace performance. In the case of the CLES check, there are different ways to measure job performance. For example, flight data could be collected during exercises which require the CLES, and analytics of the flight data could determine performance.

Act in Accordance with the Findings

The purpose of the evaluation is to ensure that the desired performance outcomes are being exhibited on the job (not just in training). Therefore, the evaluation needs to include a plan to address whatever performance issues are

uncovered; and with the ability to collect better data, we can isolate whether the performance problem is related to training or not. This action plan must identify the appropriate stakeholders, resources allocated to the evaluation of the collected data, and resources allocated to addressing issues.

The action plan should be developed by the training evaluation team comprised of stakeholders, which includes, at a minimum, a learning specialist, an organizational leader, an executive-level leader, and representatives from the community served by the training. The learning specialist will be responsible for the curriculum and making decisions about any changes that need to be made to the training based on the evaluation findings. The organizational leader will be responsible for making systems changes or addressing other non-training challenges which impede transfer. The executive leader is a critical stakeholder because it is this person who will ensure that the team has the authority and support needed to take whatever actions are needed. People who might represent the community served by the training include managers or others whose work is impacted by the performance of the people who were trained. Other stakeholders may also be included, as long as they have true stake in the outcome.

The work of the evaluation team begins before the training is actually delivered. The stakeholders will formulate the plan of action, beginning with formulating the questions they want the evaluation to answer. The group will determine what data or information they seek as well as how and when they will collect and analyze the information, and how they plan to act upon the findings. (See **Figure 5**).

Once the training has been implemented, and the various data has been collected, two important stakeholders should be added to the group: instructors and learners. Including a learner (or a few) as well as an instructor can provide insights on the data that might not otherwise be clear. If we refer back to the example of the forklift drivers, it is the drivers themselves (learners) who could best explain the reasons they had for not using what they learned when they were on the job. Further, the learner and instructor can also provide another perspective about methods for addressing any of the factors impacting training transfer.

We developed a diagram (Figure 5) depicting the ID process which incorporates this form of evaluation, and ensures the training is focused on workplace performance. The organizational problem is what triggers the need for a training analysis. Once it is determined that there is a workplace performance problem, then the training development begins.

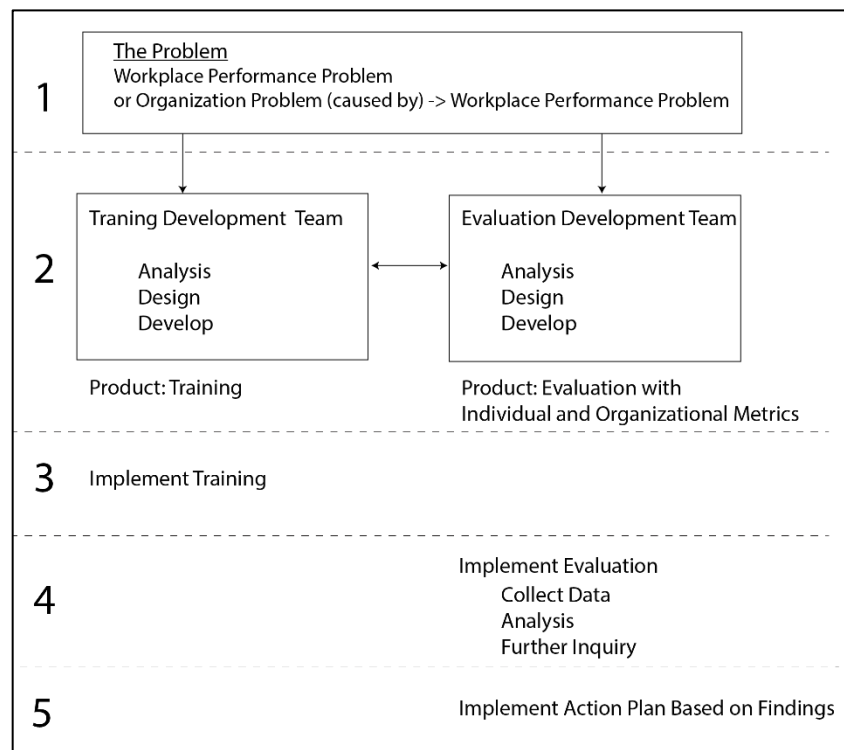


Figure 5. Instructional Design Process Focused on Workplace Performance

The training development team should be informing the efforts and approach of the evaluation, and the evaluation should inform the training (shown in phase 2). The evaluation team determines what will be measured on the job; and the training team figures out how to provide training that helps the workforce excel at those measures. The teams work together to determine how they will know if the workforce is doing what they should be and posits the areas where training transfer may be at greatest risk. The evaluation team determines what data is needed and develops a plan for collecting it. This collaborative or overlapping effort can help to prevent reliance on training for non-training problems. After the training is implemented (phase 3), the evaluation begins (phase 4). And, finally, the evaluation development team is responsible to ensure that its action plan is carried out based on the findings (phase 5). The action plan might include changes to the training, changes to the workplace procedures, an initiative designed to address manager support, or possibly a new approach to addressing the problem.

While this last stage of the evaluation may sound overwhelming, it need not be. The evaluation can be limited in scope so long as it meets the three training evaluation criteria of following a systematic method, asking (and answering) a specific set of questions, and including a plan to act upon the findings.

CONCLUSION

Evaluations enable us to determine whether, and how well, the training accomplished all of the goals and objectives that the training was developed to address (Keen & Berge, 2014). We've already noted that few organizations actually conduct evaluations of their training. There are several reasons for this, including challenges related to: "time between training and opportunity to use the skill or knowledge; difficulty of evaluating training and outcomes for complex skills or problem-solving; learning bleed – more than one training event or other events that may impact the outcome; determining whether to use qualitative or quantitative data; lack of defined goals" (Short, 2009).

Today we have methods for collecting and analyzing data that can help us to address many of the challenges that Short (2009) noted. We must not use the excuse of something being too hard or too difficult. We certainly can evaluate training for complex skills and problem solving. We can capture more data, and more meaningful data than we had in the past, allowing us to better understand how people are performing on the job. And we can link that data to training so that we can find out if training is needed, did training work, etc.

We have outlined an evaluation method that includes all three components of an evaluation: following a systematic process, seeking to answer specific questions, and having an express intent to take action based on findings. Using a research-based method to ask questions about the workplace provides information that the organization can use to figure out *why* the workforce is not performing as they were trained. Armed with information about where the breakdown is happening, the organization can do something about the problem: addressing the "why" head on. If the problem is that there is lack of support for the trained behavior, then the organization must look for ways to address the cultural problem. This may require efforts aimed at showing managers the benefit of doing things differently or a combination of approaches with people at all levels of the organization. Whatever the evaluation reveals, the response can be laser focused on addressing the actual problem which is causing a lack of training transfer.

Training is an important component in addressing complex problems facing our military and workforce. We must use the best tools available to ensure that our training is having the desired effect on workplace performance.

REFERENCES

- Advanced Distributed Learning Initiative. (2017). *Distributed learning gap report: Needs, challenges, and primary gaps impacting the implementation of distributed learning in the DoD*.
- Alvarez, K., Salas, E., & Garofano, C. M. (2004). An integrated model of training evaluation and effectiveness. *Human Resource Development Review* 3(4), 385–416. doi: 10.1177/1534484304270820
- Barnett, K. & Mattox, J.R. II (2010). Measuring success and ROI in corporate training. *Journal of Asynchronous Learning Networks* 14(2), 28–44.
- Bates, R., Holton III, E. F., & Hatala, J. P. (2012). A revised learning transfer system inventory: factorial replication and validation. *Human Resource Development International* 15(5), 549–569.

- Blume, B., Ford, J., Baldwin, T., & Huang, J. (2010). Transfer of training: A meta-analytic review. *Journal of Management* 36(4), 1065–1105.
- Bove, L. (2018). *Evaluation of compliance training in government contracting*. Unpublished manuscript, James Madison University, Harrisonburg, VA.
- Griffin, R. (2011). Perspective on practice: Seeing the wood for the trees: workplace learning evaluation. *Journal of European Industrial Training* 35(8), 841–850.
- Hannum, W. (2009a) Training Myths: False Beliefs that Limit the Efficiency and Effectiveness of Training Solutions, Part 1. *Performance Improvement*, 48(2), 26–30.
- Hannum, W. (2009b) Training Myths: False Beliefs that Limit the Efficiency and Effectiveness of Training Solutions, Part 2. *Performance Improvement*, 48(6), 25–29.
- Hannum, W. (2005). Instructional systems development: A 30-year retrospective. *Educational Technology* 45(4), 5–21.
- Haskins, M., & Shaffer, G. (2011). Assessing professional development program impact. *Strategic HR Review* 10(1), 15–20.
- Holton, E. F. (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly* 7(1), 5–21.
- Holton, E.F. (2005). Holton's evaluation model: New evidence and construct elaborations. *Advances in Developing Human Resources* 7(1), 37–54.
- Keen, C. M. & Berge, Z. L. (Nov. 2014). Beyond cost justification: Evaluation frameworks in corporate distance training. *Performance Improvement* 53(10), 22–28. DOI: 10.1002/pfi.21443
- Kirkpatrick, D. (1996). Great ideas revisited. *Training and Development*, 50(1), 54–59.
- Knowles, M. (1980). *The modern practice of adult education: From pedagogy to andragogy*. Englewood Cliffs, NJ: Prentice Hall.
- Larter, D. B. (2017, Nov. 1). Navy crews at fault in fatal collisions, investigations find. *Defense News*. Retrieved from <https://www.defensenews.com/breaking-news/2017/11/01/navy-crews-at-fault-in-fatal-collisions-investigations-find/>
- Russ-Eft, D., & Preskill, H. (2009). *Evaluation in organizations: A systematic approach to enhancing learning, performance, and change* (2nd ed.). Philadelphia: Basic Books.
- Rustici, M. (2014, August 27). *How AT&T is using xAPI to enhance compliance training*. Retrieved from <https://xapi.com/webinar-case-study-att>
- Short, T. (2009). Exploring the vacuum in training evaluation: is this a case of mission impossible? *Development and Learning in Organizations* 23(5), 15–18. DOI: <https://doi.org/10.1108/14777280910982933>
- Wang, G. G., & Wilcox, D. (2006). Training evaluation: Knowing more than is practiced. *Advances in Developing Human Resources*, 8(4), 528–539.