

MEASURING SUCCESS AND ROI IN CORPORATE TRAINING

Kent Barnett, MBA

John R. Mattox, II, Ph.D.

KnowledgeAdvisors

ABSTRACT

When measuring outcomes in corporate training, the authors recommend that it is essential to introduce a comprehensive plan, especially when resources are limited and the company needs are vast. The authors hone in on five critical components for shaping a measurement plan to determine the success and ROI of training. The plan's components should provide a roadmap to address complex corporate training environments in which large numbers of courses are delivered to thousands of learners. Recommendations offered apply equally in smaller, less complex organizations. Following a brief historical perspective covering the development of evaluation methods, the authors examine each of their five critical components—strategy, measurement models, resources, measures and cultural readiness. They claim that while their approach applies to all learning methods, it is especially useful in technology-mediated programs, such as self-paced, web-based, online-facilitated, and simulation courses.

KEYWORDS

Outcomes, Metrics, Resources, Measurement Plan, Evaluation, Cultural Readiness

I. INTRODUCTION

Today's approach to measuring success in corporate training is a complex mix of theory and practice and trial and error, with key contributions derived from evaluation theory, instructional design, technology, statistics and basic business processes. Even more intriguing is how the interaction of business, training, and technology continuously alters the way in which training is delivered and in turn influences its evaluation.

What does success mean? For some organizations, it means merely having enough data to meet compliance regulations. More robustly, others wish to determine the quality of courses across their curriculum, culling those that appear less effective. Still others try to estimate the ROI of various learning methodologies, while yet another group creates real-time dashboards to track the amount, quality and cost of training across business units. Perceptions of success vary by organization, and even within organizations, as do methods of evaluation.

II. HISTORY

Let's begin by drawing the family tree that formed the industry we know today. Training evaluation traces its genealogical roots back along two ancestral lines—at one end are academic models drawn from instructional systems design and business and military practice at the other. Early practitioners emerged from thought leaders across many disciplines who laid the groundwork for where we are now. Undoubtedly, the future will also depend on continued cross-pollination.

A. Instructional Systems Design

Intent on improving learning outcomes by creating more effective training, academics in the latter half of the last century introduced instructional systems design (ISD). According to Rothwell & Kazanas[1], “the chief aim of instructional design is to improve employee performance to increase organizational efficiency and effectiveness.” One of the most notable, and perhaps the most widely employed ISD model, consists of five basic steps: Analyze, Design, Develop, Implement, and Evaluate (ADDIE) [2, 3, 4]. In many corporations, ADDIE constitutes the framework for building training and is so ubiquitous that it is often the default approach.

For our purposes, the relevant step is the last—evaluation. Originally, Branson [2] called it “control,” reflecting the influence of Total Quality Management (TQM) as a way to identify errors and act to eliminate them. Today, “evaluation” is the term of choice because it has broader implications than merely measurement, feedback and improvement. For example, Bramley and Newby [5] propose that evaluation serves five general roles: feedback, control, research, intervention and power games.

B. Government, Military, and Business Leaders

During major military conflicts the United States faced in the past century, the armed forces sought academic participation to improve the way military personnel were trained [6]. Then, as now, competitive advantage is gained on the battlefield, not only when soldiers are properly trained, but also when they are trained faster and more efficiently than enemy soldiers. Similarly, companies need to train their employees to perform as well or better than their competitors. Becker [7] demonstrated that investments in training are worth the cost, both for individuals who pursue college degrees as well as for companies aiming to achieve ROI by training personnel.

In an example from the 1960’s, Leigh [6] shows how instructional theory contributed to educational improvement, describing how Robert Morgan proposed sweeping changes in elementary and high school curricula. Morgan’s program was called Educational Systems for the 1970’s (ES’70) and was adopted by the US Office of Education.

Morgan engaged an array of experts in the field of learning, cognition, and instructional design to contribute to the project and carried out multiple experiments in a variety of settings. Of these was Leslie Briggs, who had demonstrated that an instructionally designed course could yield up to 2:1 increase over conventionally designed courses in terms of achievement, reduction in variance, and reduction of time-to-completion—this effect was four times that of the control group which received no training.

In this example, the compelling result is not the design, but the metrics. Strikingly, it concluded that a well-designed course could yield a 2:1 improvement over other training. Leigh [6] also showed how good design principals can transfer internationally and still yield results:

In 1970, Morgan partnered with the Florida Research and Development Advisory Board to conduct a nation-wide educational reform project in South Korea. Faced with the task of increasing the achievement of learners while at the same time *reducing* the cost of schooling from \$41.27 per student per year, Morgan applied some of the same techniques as had been piloted in the ES’70 project and achieved striking results: an increase in student achievement, a more efficient organization of instructors and course content, an increased teacher to student ratio, a reduction in salary cost, and a reduction in yearly per student cost by \$9.80.

This story is just one example of the power of metrics, conveying the value of training evaluation using success metrics and ROI. Without a doubt, Morgan’s measures are those every training professional would like to have to demonstrate the effectiveness of their programs.

If these metrics sound similar to ROI metrics, they are. Return on investment is a simple mathematical

formula that balances benefits against cost. Not long after Morgan reported his conclusions, in the mid-1980s, Jack Phillips introduced ROI in corporate settings [8, 9]. Phillips' work covered improvements in supervisor effectiveness and training and employee retention. Business executives, who understood ROI as a useful business tool in decision making, applied his methods liberally in training and human capital interventions. Today, Jack and Patti Phillips, who head the ROI Institute (www.roiinstitute.com), are renowned for their ROI methodology.

The convergence between academics and practitioners is a supply-and-demand relationship. Academics supply the models to meet the demand of practitioners. A prime example is the IPISD model [2], collaboration between the US Armed Forces and Florida State University. The model is almost identical to ADDIE.

Why is the marriage between supply and demand so important in training evaluation? As the demand for faster, more effective training spearheads the evolution of corporate training, the same forces also generate change in evaluation.

To look closely at the changes, it's best to begin with the founder of modern training evaluation, Donald Kirkpatrick. In the mid-1950's, Kirkpatrick was asked to evaluate the effectiveness of a local training program. Structuring his evaluation with four levels—Reaction, Learning, Behavior, and Results [10], with Level 1 focused on learner reactions, he asked, "How did learners react to the training?" Did they like it? Would they recommend it to others? Did it meet the learning objectives and their learning needs? Level 2 determined whether learners gained knowledge and skills during training. Level 3 focused on whether training transferred to on-the-job behaviors. Finally, Level 4 investigated whether training had an impact on the bottom line. Later, his model formed the subject of his doctoral dissertation and with the publication in 1959 of a series of four articles by the American Society for Training and Development (ASTD) covering each level in detail, his model became widely popular. Its simplicity and utility ensured it would be adopted quickly and has been employed by many ever since. "E" in the ADDIE model is often thought synonymous with Kirkpatrick's four levels.

Longevity is a good indicator of the quality and usefulness of a model, and Kirkpatrick's is certainly the longest-running king. However, during the past 50 years, as the nature of training has changed, methods have had to change right along with it. Chief among the transformations is the widespread introduction of distance learning techniques—CD-ROMs, online job aids, self-paced web-based training, online facilitated classrooms, and complex simulations with avatars, among many other technologies. To keep up, evaluation tools and techniques have had to be modified. Evaluations are now delivered as online surveys; measurement software is integrated with learning management systems; and within corporate universities, scorecards produce real-time results for thousands of courses.

A model does not make a strategy and a strategy is useless unless there are resources to execute it. So how and where do measurement models fit in corporate training? In order for learning organizations to be successful at measurement, they must address five critical components:

- Develop a measurement strategy that aligns with the business
- Apply a measurement framework that fits the strategy
- Align the right resources
- Select the right measures for the organization
- Ensure the organization is culturally prepared for change

III. CRITICAL COMPONENTS FOR MEASURING CORPORATE TRAINING

To evaluate corporate training successfully, there are certain components that must be addressed. While some organizations may be able to evaluate training with less than a full complement of these elements, companies that consider all five are more likely to achieve measurement success. Although not entirely sequential, the components are presented in order of importance because decisions typically cascade from one to the next.

A. Measurement Strategy

Companies must develop a measurement strategy that aligns with the needs of the business and equally with corporate learning strategy. Why is a measurement strategy required? It is necessary because financial and personnel resources are never sufficiently abundant to measure everything. Your strategy outlines what should be measured, what framework should be used, what measures should be applied, and what resources should be used. Finally, your company must be prepared to accept and implement the results. Your strategy may not dictate whether e-learning should be measured along the same lines as instructor-led training, but it should guarantee that both will be subject to evaluation.

Boudreau and Ramstad [11] showed that strategic advantage can be gained by applying decision science to the data that businesses collect. A measurement strategy takes a long view of the nuts and bolts of gathering, processing and reporting data so that learning and business leaders will have exactly what they need to make informed decisions—timely, accurate and abundant information about where and how to apply resources.

Typically, there are three principal driving forces that spur organizations to evaluate training:

- Compliance requirements established by industry regulators,
- Information demands from learning and business leaders,
- Or a combination of both.

For some organizations, training evaluation is not optional; regulators require it. For example, in the pharmaceuticals industry, sales representatives must know their products and accurately inform physicians about drug interactions and dosage. Before representatives can engage doctors, drug companies must demonstrate that their sales force has attended training and gained product competency. Regulators often leave the evaluation process up to the pharmaceutical house itself. While some organizations distribute post-course surveys, others implement knowledge tests or introduce role-playing and performance appraisal. In other industries, such as financial services and healthcare, professionals must collect certified continuing education credits to maintain licensure. Regulators may stipulate that courses must be evaluated in order to meet continuing education requirements.

In compliance-based evaluation, regulations help learning staff set course quality standards to encourage continuous improvement. In theory, improvement cycles help ensure that learners gain what they need to know in order to succeed at their job. Unfortunately, objectives are not always met. While regulators may monitor whether evaluations were conducted, they may not seriously investigate whether results were used to improve learning.

When executives require value-based information—measures that move beyond the minimum requirements set by regulators—the strategy must adjust to focus on a much narrower set of courses. Not all courses are created equal, and since resources for evaluation can be spotty, value-based evaluations commonly focus on high-profile, high-impact, high-cost programs designed to meet strategic business needs. Clearly, for these programs, you need to know whether your training is achieving its goals and that it provides the knowledge and skills that support business initiatives in a cost-effective way. Table 1 shows a simple structure for implementing an evaluation strategy.

Strategy	Approach	Reach across the Curricula
Compliance	Mandatory evaluation of every program; Must adhere to regulators standards for evaluation; Typically apply standardized methods.	Broad—every course is evaluated
Value	Selectively apply custom evaluation theories to high-profile, high-impact, high-cost programs.	Narrow—a small number of courses are evaluated
Mixed	Apply a combined compliance- and value-based approach.	Broad to meet compliance requirements and narrow to meet information needs

Table 1. Measurement Strategies

B. Theoretical Framework

Training evaluation typically accomplishes two things: it determines the effectiveness of training and identifies areas that need to be revised. Several evaluation models exist (e.g., Kirkpatrick’s Four Levels of Evaluation) to accomplish these tasks. The most prevalent and useful models are summarized in Table 2.

Measurement Model	Author(s)	Basic Components
Experimental and Quasi-Experimental Designs	Shadish, Cook and Campbell (2002)	These models are drawn from the scientific method where an individual hypothesis is tested by controlling as many variables as possible. A clinical trial epitomizes this approach, wherein a drug is administered in varying doses to a large number of people randomly assigned to groups A, B, and C, while group D receives a placebo, and group E is a no-dose control.
4 Levels of Training Evaluation	Donald Kirkpatrick (1998)	This model measures learner opinions about Reaction, Learning, Application and Impact immediately after training.
ROI	Jack Phillips	Measure learner opinions about Reaction,

Methodology	(1997)	Learning, Application, Business Results and Return on Investment. Follow-up surveys are recommended to identify changes in behavior and business impact. Behaviors are monetized to provide a benefits value for the ROI calculation.
Learning Impact Model	Josh Bersin (2009)	Nine components comprise this model: satisfaction, learning, adoption, utility, efficiency, alignment, attainment, individual performance, and organizational performance. Unique components which differentiate this model from Kirkpatrick and Phillips are efficiency, utility and alignment. Bersin does not advocate ROI as an end in itself.
Success Case Method	Robert Brinkerhoff (2003)	This approach advocates using a small number of questions (typically 5) to gain a “pulse” among learners. The focus is not on the average scores. Rather the tails of the distributions are more interesting. Follow-up interviews are conducted with learners who provided extremely high or extremely low ratings to determine causes of success and failure.
AEIOU	Mari Kemis and David Walker (2000); Simonson (1997)	This model has five components: Accountability, Effectiveness, Impact, Outcomes and Unanticipated consequences. The unique aspect of this model is Unanticipated consequences.

Table 2. Measurement Models

The aim of these models is to discover the direct, isolated impact of learning programs on individual and business performance. They try to answer the question: Did training cause performance improvement? The ability to determine causation with validity and reliability is essential to demonstrating the value of a program. However, causation is often difficult and almost always expensive to determine. Empirically, causation is determined by employing the scientific method and an experimental design involving multiple training groups, non-training control groups, random assignment to conditions, large numbers of participants, and multiple pre- and post-training measures [12]. Quasi-experimental designs attempt to apply the same level of rigor but lack random assignment to groups, a critical requirement for experimental design. Impact studies are like pharmaceutical company clinical trials used to determine the

effectiveness of new drugs. Instead of these, alternative evaluation approaches have been developed with a more pragmatic goal in mind—to accommodate large-scale evaluations more efficiently. The desired outcome is the same: to gain information about causation with the next best thing, reliable indicators of success.

This is where measurement strategy and theoretical models intersect. Effective measurement strategy matches an appropriate theoretical model with a desired outcome. A training evaluation strategy predicated on evaluating every program with an experimental design is destined to fail because it will not be cost-effective across the curriculum, nor will it provide evaluation results in a timely manner.

Phillips' ROI Methodology [13] nicely balances both the need to determine causation and the need to control cost and effort by strategically estimating the need to apply levels of evaluation. For companies that must comply with regulations, Levels 1 and 2 may be mandatory. Consequently, resources must be allocated to maintain compliance. Additional resources may be allocated selectively to evaluate for high-profile, high-impact courses at Levels 3 through 5. In fact, Phillips recommends that most programs be evaluated at Level 1 and that very few are investigated for ROI. Not every program is worth such deep study and the cost may be prohibitive. Kirkpatrick [10] makes similar recommendations and the training industry has followed suit. ASTD has published survey results showing the extent to which learning organizations apply the Kirkpatrick and Phillips models across their curriculum. Respondents indicated that Level 1 was applied to nearly all courses. The percentages cascaded down as levels increased until reaching ROI which was applied to only 5–8% of courses [14].

When it comes to e-learning, Levels 1 and 2 on both the Kirkpatrick or Phillips models are usually sufficient. A typical one- to two-hour course usually focuses on knowledge transfer. Level 1 provides a quick check on the quality of the course and learner satisfaction with training. Level 2 determines whether training transferred requisite knowledge. More complex e-learning courses that focus on skill building and behaviors—especially simulation-based courses—are more likely to require evaluations at Levels 3, 4 and ROI.

C. Leading Indicators

Phillips suggests using post-course surveys compared with leading indicators to eliminate expensive impact studies. Leading indicators are compared to follow-up ratings by learners two to three months following training. They are validated when highly correlated with follow-up scores. Subsequently, to save time and resources, the process can be dropped. Alliger, Tannenbaum, Bennett, Traver & Shotland [15] conducted a meta-analysis of training evaluation studies and found reliable relationships across levels of evaluation.

A critical step in achieving a baseline for leading indicators is to create effective post-course evaluations that collect data about each levels of evaluation. Berk [16] plays on the industry slang of “smile sheets” for post-course evaluations, calling them “smart sheets” when they include questions that address each level. Barnett and Berk [17] have demonstrated the validity of Phillips' conclusion. Using the KnowledgeAdvisors evaluation system called, Metrics that Matter™, they collected more than 750 million data points using “smart sheets” and follow-up evaluations. Simple correlations show the strength of the relationship between leading indicators of success (post-course ratings) and actual indicators (follow-up ratings). Employing structural equation modeling, Bontis & KnowledgeAdvisors [18] also demonstrated causal links between training events, on-the-job performance, and business results.

So does such research on leading indicators obviate the need for impact studies? Not really. Some programs clearly merit the time and effort required to demonstrate that training produced a substantial and intended return.

Another evaluator worth mentioning is Robert Brinkerhoff [19] and his Success Case Method. His book, *The Success Case Method: Find Out Quickly What's Working and What's Not* emphasizes the need for efficiency and speed during evaluations. Using a survey tool, a small number of Likert scale questions are

asked of a target population, making the survey process quick and painless. When the results are examined, the average scores are deemed less important than the top and bottom extremes of the distribution. To gain insight about success and failure, the evaluation team identifies respondents who provided extremely high or low scores, and then the team conducts interviews. While efficient on the front-end, the resources required to conduct interviews can be costly and time consuming. However, the process is scalable when interviews are only conducted for a small portion of the curriculum where problems may be occurring.

Another evaluation model comes from Bersin [20] who offers the Impact Measurement Framework with nine areas of concentration, some of which are similar to both Kirkpatrick's and Phillips' models—adoption, utility, efficiency, alignment, attainment, satisfaction, learning, individual performance and organizational performance. Bersin's model is unique in that it emphasizes efficiency and highlights the need to align evaluation with learning and business objectives. Noting that learning management systems are good at collecting large quantities of useful data, efficiency analysis might review metrics such as the number of people trained, employees trained per course, hours of training per person, total cost of training, cost per learning hour, cost per employee, and repurposing, among other data. Regarding alignment, if your company is rolling out a new sales initiative, for example, your learning team should support it effectively with appropriate training and evaluation. Likewise, sufficient measurement resources should be allocated for this program for evaluation if it is a high-cost, high profile course rather than invested elsewhere.

A less well-known evaluation method, AEIOU, captures many of the same elements addressed by earlier models, but analyses measures in slightly differently by these categories—Accountability, Effectiveness, Impact, Outcomes and Unanticipated consequences [21, 22]. The unique aspect of the model is the focus on unanticipated consequences. Consider the case of managers sent to technical training to appreciate a new process employed by their direct reports. Counterproductively, they may perform the skills themselves, rather than supervise their staff. Another example might be the fact that training can lead to loss of productivity, not because employee skills decline because of training, but because they may quit, taking their newly acquired skills elsewhere, leaving the company without the talented workforce it had counted on. Simonson [21] documents how the AEIOU model has been used to evaluate distance learning education and cites several technical papers which also used this approach.

One final model is worth mentioning here, Six Sigma [23], a quality assurance method that adopts the DMAIC model in which a process is Defined, Measured, Analyzed, Improved and Controlled. Originated by Motorola, the cyclical measurement process attempts to eliminate manufacturing defects, but the model can be applied to any process, including training. In training, the DMAIC process focuses on measuring training defects (e.g., low course ratings). Because the measurement is continuous and cyclical, a subsequent measurement of the same program should show a decrease in defects—assuming that the other steps (analyze, improve, and control) were applied to improve quality.

Let's now turn to measurement strategy, how each model might usefully help the learning staff meet its measurement goals. All models, except experimental design, are sufficiently robust and practical to meet scalability requirements of large organizations. The question for each company is—which model produces the information required? Optimally, the measurement model you select and the data subsequently generated should align with your strategy (e.g., compliance vs. value vs. mixed) to provide managers with the information they need to make decisions.

IV. RESOURCES

Available resources play a key role in determining both your short- and long-term measurement strategy. Resources give you the capacity to achieve measurement goals. Three key resources to consider during

planning are financial, human and technology.

A. Financial

High-quality evaluation is difficult to accomplish on a shoestring because personnel (or consultants) with highly specialized knowledge are costly. So are systems required to scale processes in large organizations. What is a good benchmark for assessing whether too much or too little is being spent on measurement? A survey by Ray [24] indicates most organizations spend less than 4% of their learning and development budgets on evaluation and metrics. Of those, 59% spend less than 1% on measurement. Similarly, Bersin [25] found that 39% of organizations spend less than 1% of their training budget on measurement; 94.3% spend less than 5%.

B. Human Resources

Once a budget is established, human resources are needed, among other things, to set the strategy, administer the system, and build capacity to exploit the data. Finding the right mix of internal and external resources can be tenuous especially during the first year the measurement group is established. Luckily, resources required now are not as great as it was needed 10 years ago. Mattox, Jinkerson & Hanssen [26] examined the dramatic changes in the size of learning and development measurement groups across three professional services firms. They found that advances in technology, especially with the widespread introduction of learning management and measurement systems, now allow more modest groups to accomplish the same goals that earlier required much larger staff.

Dewey, Montrosse, Schroter, Sullins, and Mattox [27] documented competencies required of professional evaluators by validating whether what is taught to evaluators is actually what is sought by employers. Notably, once hired, professional evaluators serve as the content experts for the organization, but they cannot fly solo. Their expertise alone is not sufficient to make the group successful in large organizations. Your measurement team should have access to or be staffed with personnel who know how to implement and use technology systems that can make the group more efficient. Additional quantitative and qualitative data analysts are also needed to help translate data into useful information.

C. Systems

Systems are required to make the measurement process scalable, especially for large organizations with thousands of learners, thousands of courses, and multiple national and international locations. In measurement, Barnett and Berk [17] suggest that scalable systems cover four practical data needs: collection, storage, processing and reporting. Efficient systems use standardized forms for data collection, helping to create robust reporting benchmarks. It's best to store data in a central repository with sufficient security to ensure safety but enough flexibility to allow data to be mined as needed. In processing data, it's wise to leverage the system to perform complex filtering, queries, and analysis with minimal human intervention. Your reports should present data with tabular and graphical results using standards report software and additional reports should be available easily with an ad hoc query. Your objective is to maximize the capabilities of the system, turning individual data points into productive information.

System resources must meet the needs of the business. If you need only know how many employees attended your e-learning curriculum and the re-use ratio, your learning management system will suffice. If your strategy dictates that Kirkpatrick's Levels 1 and 2 evaluations are mandatory to meet compliance regulations, then your measurement system must accommodate those needs. Integration with your learning management system is essential, especially for e-learning, typically available online only, leaving little or no opportunity to reach dispersed learners after training is completed except virtually.

V. MEASURES

While training evaluation models are plentiful, measures seem unlimited. Rossi & Freeman [28] divide evaluation metrics into two camps: process and impact. Process measures are those that address training

operations. How many people registered for an e-learning product? How many launched it? How many completed it? What was the development cost? What is the re-use ratio? Impact measures address the effectiveness of training. Did attendees learn new knowledge and skills? Will training improve job performance? Will performance improvement affect the bottom line? Both types of measures are important, but there is clearly more value associated with the latter than the former. The distinction between these measures also helps us consider cause and effect. Process measures monitor causes (e.g., differing dosages of training) and impact measures describe the effect.

In the end, actual measures used within an organization are determined by the strategy and model. As an example, let's look at Boudreau and Ramstead [11] who offer the HR BRidge Framework with a focus on three important groups of measures: efficiency, effectiveness and outcomes. These measures arose out of research relating to human capital metrics and apply nicely within the learning measurement space. Examples of all three types are given in Figure 1. Efficiency measures address whether the investment was high or low, whether enough learners attended training, and whether learning is actively pursued. Effectiveness measures assess the quality of training, whether it affects job performance, and whether that performance realized is a valuable outcome compared with the investment. Finally, outcome measures focus on the productivity of trainees, resulting revenue produced (or not produced), and whether cost reductions can be achieved through training. The intersection of all three is a theoretical point of optimization in which the corporation's investment in training is at just the right level to achieve the optimum amount of output from employees.

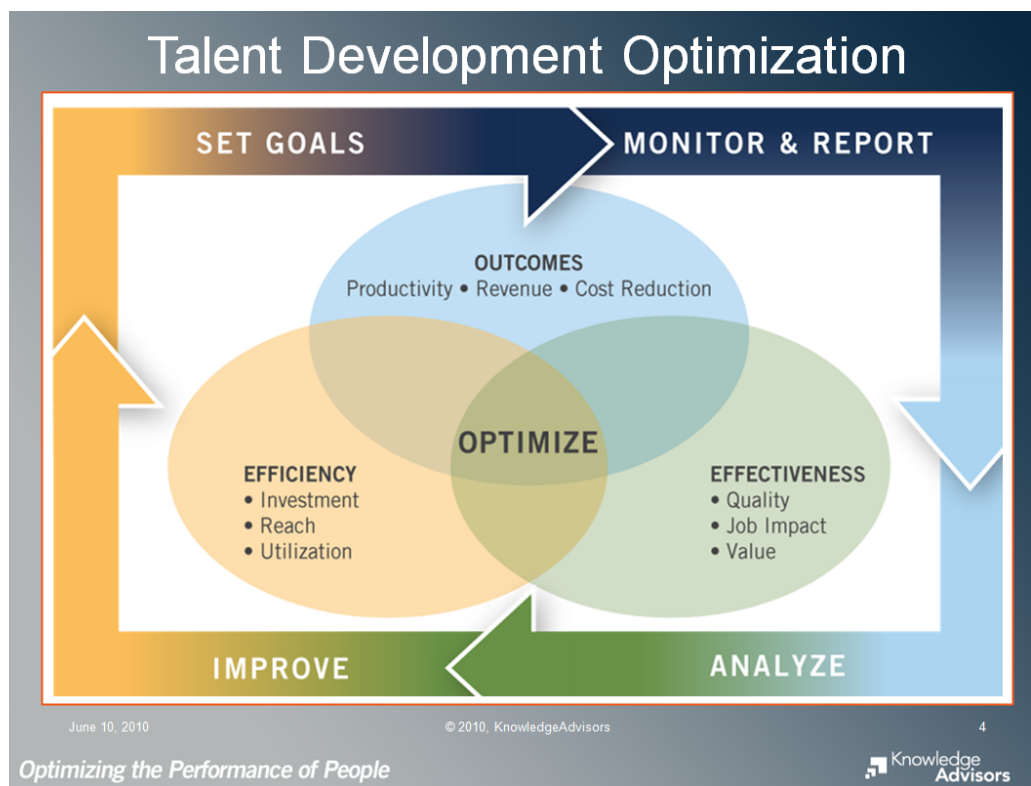


Figure 1. Optimizing Performance (Adapted from Boudreau & Ramstead, 2007)

Regardless of the strategy and measurement model employed, a variety of questions can be raised for e-learning courses to help demonstrate effectiveness. Table 3 contains examples of valuable questions to

determine the quality of many e-learning programs. Distinctions are made for self-paced web courses, on-line facilitated courses and simulations. The suggested scale for these questions ranges from strongly disagree to strongly agree. These measures all fall within Kirkpatrick’s Level 1. Why only Level 1? Because training, not the training method, is the intervention that is intended to yield gains in knowledge, skills, and productivity. E-learning as a training method is a process measure. As such, questions related to the training process will help determine the quality of training and eventually its effectiveness. The impact of an e-learning course is measured the same way it is measured for any other course.

Question	Self-Paced Web-based	On-line Facilitated	Simulations
Course registration information was accurate and useful.	X	X	X
I was able to easily launch the training course.	X	X	X
The course materials helped me learn.	X	X	X
The pace of the course was appropriate for the material.	X	X	X
The user interface was easy to use.	X	X	X
The delivery method was appropriate for the content and objectives of the course.	X	X	X
The exercises were well-suited to the delivery method.	X	X	X
The exercises helped me learn valuable knowledge and skills.	X	X	X
Overall, the instructor was effective.		X	
The instructor’s presentation style was effective.		X	
The instructor answered questions clearly and completely.		X	
Instructions for the simulation were clear.			X

The simulation helped me learn knowledge and skills in a way that I could not have learned easily on the job.			X
---	--	--	---

Table 3. Recommended evaluation questions for e-learning courses

In addition to these metrics, it is also valuable to consider linking them to process measures from your learning management system, including such data as the number of people who attended, re-use ratio, and the cost of training. Combining these data sets helps understand efficiency and alignment [20].

Once the appropriate measures have been selected, your measurement team should focus on reporting what you uncovered to stakeholders. Patton [29] indicates that utility—understanding and using results—is the ultimate objective of any evaluation. If stakeholders are not going to use the data once it is collected, it should not be collected. Patton’s perspective should guide your measurement team. To increase the use of data by company stakeholders, here are three recommendations:

- Make your reporting process as simple as possible for your users by automating reports for delivery to end-users.
- Display graphs and tables summarizing data into useful information.
- Build dynamic scorecards aggregating results across courses and curricula, giving a wide view of your data.

VI. CULTURAL READINESS WITHIN YOUR ORGANIZATION

While cultural readiness is the last of the five components discussed here, it is the one that can most easily derail all your prior efforts. Readiness doesn’t happen. It must be suggested, nurtured and sustained among three key stakeholder groups: leaders, learning and development managers, and learners.

Leaders must know what they want, why they want it, and champion it when times are tough and measurement is on the chopping block. If they are not ready for a new measurement strategy, the initiative is likely to fail. Luckily, leaders typically sponsor these initiatives and often win over other leaders to their perspective. Learning and development teams need to know how to interpret and use the data they gather from their measurement systems. Equally important, senior managers must be ready to implement the changes they will be facing when it comes to attending training, providing feedback, taking tests, and completing follow-up evaluations. This includes learners’ supervisors, too, who may be contacted to provide feedback about performance on the job.

Evaluation Capacity Building (ECB) is “the extent to which an organization has the necessary resources and motivation to conduct, analyze, and use evaluations” [30]. Hallie Preskill, a thought leader on the topic, has published a multidisciplinary model of ECB [31] which conceptualizes evaluation knowledge, skills and attitudes required to produce a sustainable evaluation practice while accounting for the influence of leadership, culture, systems and structures, and communication. By implementing this model, your company should be able to educate and acculturate stakeholders.

Every effort to build evaluation capacity within an organization is a journey of a thousand steps. For some organizations, it’s a journey of a thousand miles—or more! There are several ways to track progress. Certainly, the measurement group can measure its own performance against its goals, but there are also external frameworks for comparison. Wettstein & Kueng [32] offer a four-stage maturity model for performance measurement systems in general.

Figure 2 shows KnowledgeAdvisors’ Measurement Maturity Model that applies to learning analytics as well as to broader human capital analytics. Additionally, KnowledgeAdvisors offers a companion assessment tool allowing companies to determine where they stand on the maturity curve and how they can improve their standing.



Figure 2. KnowledgeAdvisors’ Measurement Maturity Model

VII. EXAMPLE

At this point it seems worthwhile to share a case in which an organization successfully transformed its compliance-focused measurement processes into a robust measurement approach. Their measurement efforts helped the learning and development team determine the quality of its programs as well as identify those that needed improvement. The end result was that the organization gained recognition by *Training Magazine* as a Top 10 corporate university four years in a row.

The organization is a professional services firm that trains more than 25,000 people annually with a library that exceeds 1,000 online and instructor-led courses. Employees are dispersed throughout the country, making e-learning an efficient and effective tool for training. Table 4 describes how the organization addressed each of the five critical components discussed in this article. Table 5 displays how much of each learning methodology is evaluated on each of Phillips’ 5 levels. Superscripts indicate which aspects of evaluation are required by regulators, driven by business leaders or pursued for value.

Critical Components	How did the organization address the component?
Develop a Measurement Strategy	The original strategy was strictly compliance-focused because of industry regulations. L&D leaders reexamined their information needs and determined that a compliance and value strategy was required.
Apply a Measurement Framework	Phillips’ ROI Methodology was chosen as the core approach for measurement. Levels 1 and 2 were pursued extensively and Levels 3 – ROI were pursued selectively. As needed, other models were applied for a small number of courses.
Align Resources	The organization hired one full-time measurement expert during year one and eventually grew to include three more people. All had at least a master’s degree and experience with educational or psychological

measurement. To augment the team, financial resources were appropriated to hire external consultants as advisors and for special projects. Financial resources were also used to acquire an evaluation system and a testing system. Systems personnel were available within the L&D group to assist with integration with the learning management system.

Select Measures	Standard evaluation questions were selected and applied to all forms across learning methodologies. A few unique questions were added to certain learning methodologies. For example, self-paced web-based forms included an item about the user interface and the delivery method. Also, the instructor questions were dropped. Instructor items were retained for online facilitated courses. After implementing the measurement systems and gathering a year’s worth of data, the measurement team focused on developing a quarterly scorecard with only a handful of measures, one from each of Phillips’ 5 levels of evaluation.
Develop Cultural Readiness	Leaders were on board from the start, but like the learning and development managers, they needed some basic education about measurement models, systems and reports. The measurement team developed online learning models, held workshops, and tutored individuals upon request. Two groups served as champions for the measurement processes: a governance group comprised of directors and a manager group comprised of system / process super-users. The governance group kept leaders informed and enforced the standards with the super-users. The last group of stakeholders, learners, received many communications about pending changes in the evaluation process.

Table 4. Case Study: Addressing the Five Critical Components of Measuring Corporate Training

Level	Instructor-Led	e-Learning	Conferences
Reaction and Planned Action	100% ^a	100% ^a	100% ^a
Learning	<75% ^b	100% ^a	<50% ^b
Job Application	<5% ^c	<2% ^c	<2% ^c
Business Results	<5% ^c	<2% ^c	<2% ^c
Return on Investment	<5% ^c	<2% ^c	<2% ^c

Table 5. Amount of Measurement by Learning Methodology and Phillips’ Levels of Evaluation

^a required by regulators

^b driven by business leaders

^c pursued for value

VIII. CONCLUSION

Measuring the success and ROI of corporate training is a matter of establishing and executing a solid

strategy. Measures, including ROI, are determined by the company's learning and business needs. Success is achieved at two levels. First, when the five critical steps are accomplished to build a robust and sustainable measurement process, and second, when actual metrics begin returning valuable information about the success (or lack of success) of a program.

Moving beyond training, there are many measurement issues that will interest learning and business leaders. As examples, what factors beyond training, such as manager engagement, electronic performance support systems (EPSS), and informal learning, make training more effective? Early investigations into manager engagement indicate that it is strong lever for extending the effectiveness of training and reducing scrap learning [33]. Examples of the influence of EPSSs have been documented [34] and investigations have just begun on the effects of informal learning [35].

Value-based measurement is also gaining traction in human resources departments as organizations closely follow human capital metrics. Whether it is employee loyalty, speed to competency, or turnover, measurement plays an essential role in determining the effectiveness of an organization. Studies [36, 37] demonstrate that companies with mature and effective human capital processes contribute substantially to the bottom line. In fact, they outperform the S&P 500.

Today, the challenge may be building a successful measurement process. Tomorrow, it may expand to integrating data across human capital systems. There appears to be an endless need for measurement to help business leaders make decisions. To paraphrase Edward Hubbard, "training is either at the table working with senior management and adding value, or it is on the table perceived as a cost center that is going to get cut." By following the five critical components covered in this article, learning departments can grab a seat at the table and avoid the chopping block because success measures will meet business information needs and demonstrate the value of learning.

IX. ABOUT THE AUTHORS

Kent Barnett, chairman and CEO of KnowledgeAdvisors, the world's largest provider of learning and talent measurement solutions, was co-founder and former president of Productivity Point International (PPI). Barnett was responsible for substantial growth at PPI before it was acquired by Knowledge Universe.

John R. Mattox, II, director of research at KnowledgeAdvisors, employs advanced statistics when mining the company's database to gain insight for its roster of clients. Earlier, he led training evaluation teams at several professional services firms.

KnowledgeAdvisors
222 S. Riverside Plaza, Suite 2050
Chicago, IL 60606

X. REFERENCES

1. **Rothwell, W.J. & Kazanas, H.C.**, *Mastering the instructional design process: A systematic approach*. 2nd ed. San Francisco, CA: Jossey-Bass Publishers, 1998.
2. **Branson, R. K.**, "The interservice procedures for instructional systems development," *Educational Technology*, pp. 11-14, Mar 1978.
3. **Dick, W. & Carey, L.**, *The systematic design of instruction*, 4th ed. New York: Harper Collins Publishing, 1996.
4. **Molenda, M.**, In search of the elusive ADDIE model. Retrieved on June 9, 2010 from <http://www.indiana.edu/~molpage/In%20Search%20of%20Elusive%20ADDIE.pdf#search=%22ADDIE%20Model%20%2Bhistory%22>.
5. **Bramley, P. & Newby, A. C.**, "The evaluation of training part I: Clarifying the concept," *Journal of European & Industrial Training*, 8:6, pp. 10-16, 1984.

6. **Leigh, D.**, “A brief history of instructional design.” Retrieved June 9, 2010 from <http://www.pignc-isp.com/articles/education/brief%20history.htm>
7. **Becker, G.S.**, *Human capital: A theoretical and empirical analysis with special reference to education*, 3rd ed. Chicago, IL: University of Chicago Press, 1993.
8. **Phillips, J.J.**, *Improving supervisors’ effectiveness*, San Francisco, CA: Jossey-Bass Publishers, 1985.
9. **Phillips, J.J.**, *Recruiting, training and retaining new employees: Managing the transition from college to work*, San Francisco, CA: Jossey-Bass Publishers, 1987.
10. **Kirkpatrick, D.L.**, *Evaluating training programs: The four levels*, 2nd ed. San Francisco, CA: Berrett-Koehler Publishers, Inc., 1998.
11. **Boudreau, J.W. & Ramstad, P.M.**, *Beyond HR: The new science of human capital*, Boston, MA: Harvard Business School Press, 2007.
12. **Shadish, W.R., Cook, T.D., & Campbell, D.T.**, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin, 2002.
13. **Phillips, J.J.**, *Return on investment in training and performance improvement programs: A step-by-step manual for calculating the financial return*, Houston, TX: Gulf Publishing Company, 1997.
14. **Sugrue, B. & Riviera, J.**, *State of the industry: ASTD’s annual review of trends in workplace learning and performance*, Alexandria, VA: American Society for Training and Development, 2005.
15. **Alliger, G. M., Tannenbaum, S. I., Bennett, W. Jr., Traver, H., & Shotland, A.**, “A meta-analysis of the relations among training criteria,” *Personnel Psychology*, 50, pp. 341-358, 1997.
16. **Berk, J.**, “Emerging issues in measurement strategy,” *Chief Learning Officer*, November, pp. 34-39, 2009.
17. **Barnett, K. & Berk, J.**, *Human capital analytics: Measuring and improving learning and talent impact*, Tarentum, PA: Word Association Publishers, 2009.
18. **Bontis, N. & KnowledgeAdvisors, Inc.**, *The predictive learning impact model*, Chicago, IL: KnowledgeAdvisors, 2009.
19. **Brinkerhoff, R.**, *The success case method: Find out quickly what’s working and what’s not*, San Francisco, CA: Berrett-Koehler Publishers, Inc., 2003.
20. **Bersin, J.**, *The training measurement book: Best practices, proven methodologies and practical approaches*, San Francisco, CA: John Wiley & Sons, 2008.
21. **Simonson, M.R.**, “Evaluating teaching and learning at a distance,” *New Directions for Teaching and Learning*, 71, pp. 87-94, 1997.
22. **Kemis, M. and Walker, D.A.**, “The a-e-i-o-u approach to program evaluation,” *Journal of College Student Development*, (41:1), pp. 119-22, 2000.
23. **Pyzdek, T. & Keller, P.A.**, *The six sigma handbook: a complete guide for green belts, black belts and managers at all levels*, 3rd ed. New York, NY: McGraw-Hill, 2009.
24. **Ray, R.L.**, “The strategic impact of learning: how corporate learning strategies are driving business results,” ROI Institute’s Capturing Value For Money 12th Annual Global ROI Conference October, 2008, Dublin, Ireland. Retrieved June 9, 2010 from <http://www.impact-measurement-centre.com/LinkClick.aspx?link=Documents%2FRebecca+Ray.pdf&tabid=183&mid=616>.
25. **Bersin, J.**, High-Impact learning measurement state of the market and executive summary, Oakland, CA: Bersin & Associates, 2010.
26. **Mattox, J.R., II, Jinkerson, D.L. & Hanssen, C.E.**, “Technology drives restructuring of measurement teams in learning organizations: Doing more with less in the professional services industry.” Presentation made at the annual American Evaluation Association Conference 2007: Baltimore, MD, 2007.
27. **Dewey, J.D., Montrosse, B.E., Schroter, D.C., Sullins, C.D., & Mattox, J.R., II**, “Evaluator competencies: What’s taught versus what’s sought,” *American Journal of Evaluation*, 29:3, pp. 268-287, 2008.

28. **Rossi, P.H. & Freeman, H.E.**, *Evaluation: A systematic approach*, 5th ed. Newbury Park, CA: Sage Publications, 1993.
29. **Patton, M.Q.**, *Utilization-focused evaluations: The new century text*, 3rd ed. Thousand Oaks, CA: Sage, 1997.
30. **Gibbs, D., Napp, D., Jolly, D., Westover, B., & Uhl, G.** “Increasing evaluation capacity within community based HIV prevention programs,” *Evaluation and Program Planning*, 25, pp. 261-269, 2002.
31. **Preskill, H. & Boyle, S.**, “A multidisciplinary model of evaluation capacity building,” *American Journal of Evaluation*, 29:4, pp. 443-459, 2008.
32. **Wettstein, T. & Kueng, P.**, “A maturity model for performance measurement systems,” *Management Information Systems*, Southampton, UK: Witt Press, 2002.
33. **Brinkerhoff, R.**, “Manager engagement and training impact.” Paper presented at KnowledgeAdvisors 8th Annual Analytics Symposium. National Harbor, MD, Mar. 2010.
34. **Jury, T. and Reeves, T.**, “An EPSS for instructional design: NCR’s quality information products process,” *Design approaches and tools in education and training*, Dordrecht, The Netherlands: Kluwer Academic Publishers, 1999.
35. **Parskey, P.**, “Informal learning: Why we should embrace it, fund it, and measure it.” Paper presented at KnowledgeAdvisors 8th Annual Analytics Symposium. National Harbor, MD, 2010.
36. **Bassi, L., Harrison, P., Ludwig, J., & McMurrer, D.**, “The impact of U.S. Firms’ investments in human capital on stock prices.” Technical paper published by Bassi Investments, Inc. Retrieved June 9, 2010 , from http://www.bassi-investments.com/downloads/ResearchPaper_June2004.pdf
37. **Bassi, L. & McMurrer, D.**, “Human capital management predicts stock prices.” Technical paper. published by McBassi and Company. Retrieved June 9, 2010 from <http://www.mcbassi-company.com/documents/HCMPredictsStockPrices.pdf>.