

Kirkpatrick Model

If you're not familiar with the Kirkpatrick four levels of training evaluation, take a few moments to read the following excerpt from a research report I wrote for the eLearning Guild several years ago. It is based on my doctoral dissertation: "Usage and Value of the Kirkpatrick Four Levels of Training Evaluation." One of the great honors of my career was conducting a master dialogue with Dr. Kirkpatrick himself at an eLearning Guild conference in 2006.

In addition, I refer you to a more recent post on the subject, which picks up on the recommendations I made in the conclusion of my research: "begin with the end in mind." The post is from Don Clark's Performance Juxtaposition website.

"Introduction to Usage and Value of the Kirkpatrick Four Levels of Training Evaluation" **by Joe Pulichino, EdD.**

There is an old adage commonly cited in management circles, "If you can't measure it, you can't manage it." Most executives would likely agree that managing their organization's training function is an essential responsibility of the enterprise. Therefore, measuring and evaluating the effectiveness of that function must be a priority of the professionals charged with this responsibility. Yet, a major challenge faces these professionals: how best to perform such measurement and evaluation and report the results in a timely, cost effective, and useful manner. Where can they find a method or system to address this challenge?

Many training professionals turn to the Kirkpatrick four levels simply because it has become an industry standard for evaluating training programs over the course of forty-seven years in the literature. First described by Donald Kirkpatrick in 1959, this standard provides a simple taxonomy comprising four criteria of evaluation (Kirkpatrick originally called them steps or segments, but over the years they have become known as levels). The nature of the four-level taxonomy suggests that each level after the first succeeds from the prior level. The first level measures the student's reaction to the training and the second level what the student learned. The third level measures change in on-the-job behavior due to the training, and the fourth results in terms of specific business and financial goals and objectives for the organization.

Despite its status as the industry standard, many studies, including one conducted by Kirkpatrick himself, have shown that the full taxonomy is not widely used past the first two levels. This pattern of usage means that training practitioners might not be fully measuring, and therefore effectively managing, the impact that training and development has on two of the most important reasons for funding and providing resources for training in the first place: improvements in workplace performance and positive business results.

As a result, several important questions come up. Why are not all the levels of the taxonomy as described by Kirkpatrick used more widely by training professionals? If the measurement of training is a critical task, and the industry boasts of a standard for evaluation that is almost fifty years old, then why does so much important measurement remain undone? Certainly, one reason is that measurement of each succeeding level is more complex, more time consuming, and therefore more expensive. Kirkpatrick also posits that another reason is the lack of expertise among practitioners to conduct these higher levels of evaluation. These

reasons would appear to be valid obstacles, and they may be more difficult to overcome than others may, but the problem of limited usage of the “higher” Levels 3 and 4 may lie deeper within the nature and structure of the four levels themselves.

According to several critics, beginning with John Newstrom in 1978, there are significant problems with the Kirkpatrick four levels and with the ways that practitioners misapply them due to untested and invalid assumptions often made about them. The most notable and consequential criticism is that the four levels are a simple taxonomy, and not a full researchable model as many in the industry have come to assume (Newstrom 1978, Alliger & Janak 1989, Newstrom 1994, and Holton 1996). Ironically, this criticism is one that Kirkpatrick himself fully accepts, but does not address directly.

Nonetheless, many organizations do use the Kirkpatrick four levels and derive great value from that usage, especially Levels 3 and 4. One way of addressing the two key problems of the Kirkpatrick four levels, limited usage in the higher levels and assumptions that it can function as a model rather than a taxonomy, would be to examine the characteristics and practices of those organizations that have wrestled with the challenges of implementing Levels 3 and 4, often with positive effect on business objectives such as achieving competitive advantage or building a knowledgeable and skilled workforce.

Summary of Conclusions: Usage and Value of Kirkpatrick Evaluation

The following represents a brief summary of the conclusions I have reached as a result of this study:

1. The current usage of Kirkpatrick’s four levels continues to be widespread and indicates that it remains the industry standard for evaluating training programs.
2. The Kirkpatrick four levels are a taxonomy and not a model, but this taxonomy can be used to build a researchable model.
3. The fact that Levels 3 and 4 continue to be used less frequently has more to do with the influence of business and organizational drivers as factors in the usage decision, rather than a failure of the industry to advance the state of training evaluation beyond Kirkpatrick’s original formulation. When an organization has a good business reason to evaluate at Levels 3 or 4, they tend to do so, and they do so successfully.
4. Practitioners who use Levels 3 and 4 do so for reasons that have more to do with measuring specific Level 3 and 4 outcomes rather than to justify the training department or to determine whether to continue or discontinue programs, as was first recommended by Kirkpatrick.
5. Practitioners who use Levels 3 and 4 overcome many of the challenges cited by those who do not use Levels 3 and 4 as the main reasons that they do not.
6. Practitioners who use Levels 3 and 4 tend to obtain valuable data.
7. The value of the data tends to increase when practitioners examine relationships between outcomes among the levels and consider the impact of intervening variables on the outcomes.
8. Usage of the Kirkpatrick’s four level taxonomy is most valuable when the desired outcomes of Level 3 and 4 evaluations are identified before designing a training program and are used to guide and inform that design.

Implications and Recommendations

In both the review of the literature and the data analysis of the results, this study provides several indications that the Kirkpatrick four levels are a taxonomy, not a model. It appears that Holton and other critics were correct and responsible in pointing out this distinction as it applies to the four levels. Despite Kirkpatrick's claim that it does not make a difference if they are a model or a taxonomy, it does matter in how the four levels are applied and what results they achieve. The findings presented in this study indicate that those who make this distinction and use the four levels as a taxonomy might be more capable of building an evaluation model that is useful and valuable for measuring the effectiveness of training and the impact that training has on employee behavior and business results.

The value obtained by these practitioners appears to be related to certain best practices, namely, (a) measuring the relationship of outcomes among the four levels and (b) considering the influence intervening variables have on those outcomes. In other words, they do exactly what Kirkpatrick and his critics both say must be done. So by acknowledging, even implicitly, that Kirkpatrick's four levels are a taxonomy, practitioners have the basis and the incentive to supply those components of a researchable model that are missing in Kirkpatrick's formulation, and in doing make they make their Level 3 and 4 evaluation efforts more useful and valuable.

Therefore, the findings presented in this study have several implications for those who are charged with the responsibility of evaluating the effectiveness of training programs. Whether these practitioners are already using the Kirkpatrick four levels, especially Levels 3 and 4, or are considering using them, the data shows that evaluators can maximize their potential for successful evaluation by following a few simple guidelines drawn from this study. Even those who stand on the periphery of their organization's evaluation efforts, or have a business interest in the outcomes, can benefit from an understanding of these guidelines. These guidelines are as follows:

9. Consider the external pressures on the organization that are influencing training expenditures. When deciding whether or not to proceed with Level 3 or 4 evaluation, evaluators should consider that the size of their organization's annual training budget may not matter as much as external factors in the organization's environment that are influencing training expenditures. Regardless of the amount of money being spent on employee development, if external business pressures are causing increases in the level of expenditure, an important dynamic will be at play. Namely, training programs will be expected by management to focus on producing outcomes that will address the external pressures, largely by affecting change or improvement in employee performance and achievement of business results. In the case of this study, competitive pressures and the need to maintain a knowledgeable and skilled workforce were examined, but certainly other external factors can and should be considered. By understanding the nature of these influences more clearly, evaluation practitioners can more accurately assess the business need and organizational readiness of Level 3 and 4 evaluations. This will also help the training department understand the business issues that must be addressed and that potentially can be affected positively by the right training program.
10. Confirm that Level 3 and 4 evaluation efforts will focus on the Level 3 and 4 criteria outcomes as well as the effectiveness of training programs. With this understanding in place, the reasons for conducting a Level 3 or 4 evaluation become more centered on Level 3 or 4 criteria outcomes as opposed to justifying the existence of the training department or its budget. Therefore, it is more

likely that these evaluations would be conducted for the right reason and therefore generate the support needed to overcome the challenges of conducting them.

11. Prepare to overcome the likely reasons not to conduct Level 3 and 4 evaluations, and know that they can be overcome. Having identified good reasons to conduct an evaluation helps to overcome the reasons not to conduct one, but even if there are good reasons such as those noted above, the negatives must still be dealt with. The most significant obstacles are access to data and available time. Knowing this will focus the evaluator on enablers, such as process and technology, to find solutions to address these problems. Level 3 and 4 evaluations may be difficult to do, but this study shows that they can be done and to good purpose.
12. Plan to consider the effect of intervening variables on both behavior and results. The value of data obtained in Level 3 and 4 evaluations increases as a result of considering the effect of intervening variables. Without a thorough understanding of the influence of these variables, it becomes difficult to isolate and see the influence of training. The evaluator who sees the influence of training in the context of other variables will have a better chance of using Level 3 and 4 data to measure the effectiveness of training and, most importantly according to this study, improve training programs.
13. Plan to measure relationships among the outcomes of all levels of evaluation. The value of data obtained in Level 3 and 4 evaluations increases as a result of measuring relationships among levels. Upon determining that the organizations should proceed with Level 3 or 4 evaluations, evaluators should design the evaluation process to include measurement of the relationship of outcomes among all the levels being used, especially between Levels 2 and 3, as well as Levels 3 and 4.
14. Have confidence that the data obtained will likely be valuable to the training department and to the organization. Kirkpatrick himself is a proponent of doing something rather than nothing. “Just get started,” he admonishes practitioners. The findings presented here show that organizations who use Levels 3 and 4 report very high value levels for the data they obtain. With the knowledge that others who have gone down this path have gotten valuable results, evaluators should move forward in their efforts confident that they can produce similar value for their organizations, if they follow the lead of those who have been successful.
15. One of the intriguing actions suggested by the findings of this study is that evaluators should begin the design and planning of their evaluation program with Level 4 and work their way back to Level 1. This may perhaps be the most important implication of the study. Kirkpatrick insists that practitioners start their evaluation efforts with Level 1 and work their way step by step, or level by level, to the next succeeding level, until they are conducting Level 4 evaluations. This is only true as a matter of expediency with regard to evaluation capabilities and competence, but when it comes to design and planning a whole system or model for evaluation, the findings suggest that it might be best to start with Level 4 and work backwards. This guideline will be discussed in more detail in the paragraphs below.

By focusing on the ultimate desired outcome of the funding organization, evaluators can better accommodate identification and consideration of intervening variables. By knowing the desired end result first, evaluators would be better able to estimate the impact that the outcomes from one level will have on the next. In other words, if one knows the result, one is better able to know the behavior that will produce the result. If one knows the behavior, one is better able to know what learning needs to happen to change or improve the behavior. If one knows what needs to be learned, one is better able to know how to plan for a positive reaction by the student to the training. This sequence of design and planning, taking the four levels in reverse

order, might very well be the key to creating a researchable model out of the taxonomy. Once accomplished, such a design would then make it possible to conduct the evaluation just as Kirkpatrick suggests, from reaction on through to results. Therefore, it is suggested that before getting started with Level 1, before even designing the training program, practitioners should proceed according to the following steps:

Step 1. Understand the desired results quantitatively and qualitatively. Consider all the external and internal variables that could affect the outcome of the results. Consider how training as one variable could affect the outcome of the results. Identify the metrics and measurements that could be taken to determine whether training has an impact on results, but also that could be related back to outcomes measured at lower levels in the taxonomy.

Step 2. Understand what behaviors would have a positive or negative affect on the desired results. Consider all the variables that might affect a change or improvement in behavior. Consider how training, as one variable, could affect the behavior. Identify the metrics and measurements that could be taken to determine whether training has an impact on behavior, but also that could be related back to outcomes measured at lower levels in the taxonomy.

Step 3. Identify what needs to be learned in order to change or improve behavior and produce the desired results. Design transfer steps between levels that will more tightly bind relationships of outcome among the levels.

Step 4. Identify the desired positive reactions of the student as related to what needs to be learned, what behavior needs to be changed or improved and what results need to be achieved.

Step 5. Then, after the training has been delivered, evaluate its effectiveness by following Kirkpatrick's four steps as prescribed.

This process, which seems to be supported by these findings, would allow evaluators to do a more credible and thorough job of designing and delivering the right training programs. This process would also show that training programs have been effective at all levels of evaluation.

Recommendations for Further Research

This study has been deliberately designed to support and encourage additional research on the usage and value of Kirkpatrick's four levels. It is suggested that this additional research take two forms: (a) new, yet similar descriptive studies that survey other groups of training professionals and evaluation practitioners on the topic; and (b) case studies of organizations who are using Kirkpatrick Levels 3 and 4, especially among those who report that they obtain valuable data, measure relationships among levels, and consider intervening variables. The intention of these two lines of inquiry would be to continue monitoring trends, to expand the number and kind of variables under investigation, and to begin examining in greater detail the best practices of successful evaluation efforts.

New descriptive studies. In terms of data collection, the format of this study's survey instrument proved quite effective for moving participants through the survey and directing them to the appropriate questions in a logical fashion. The question types and sequencing of questions could be easily reused and, with appropriate changes to the content, could greatly expand the scope of examination while retaining a common ground between this study and any new research undertakings. It is suggested that even with the introduction and

addition of other question types, the focus of these new descriptive studies could and should include the following:

- Additional external pressures that influence training expenditures
- Detailed expenditures and budget levels for training and for evaluation efforts
- Circumstances under which the four levels are used
- Differences in usage by industry and demographics of the organization
- Additional reasons why the four levels are used, and why they are not
- Value of the data in terms of additional measurement criteria
- Correlations of four level outcomes among other data points
- Consideration of additional intervening variables

Such studies could provide a wealth of new data that would enable evaluation practitioners to stay current with industry trends and find clues and indicators of successful best practices. These studies may also identify specific evaluation efforts worthy of greater scrutiny in the form of cases studies.

Case studies of best practices. By using the findings of these suggested descriptive studies as guides, evaluators could then examine and report on case studies of actual evaluation efforts. In doing so, they would be able to access stories and examples of real success. While it is valuable on one level to have survey participants report that they have overcome challenges, measure relationships among levels, consider intervening variables, and obtain valuable data, that value would be enhanced through discovery and exploration of exactly what they do and how they do it. There are likely many buried treasures in the stories of those evaluators whose organizations use Levels 3 and 4 consistently and successfully. These organizations should be examined and queried thoroughly so that the industry—and especially the community of evaluation practitioners—can develop a more comprehensive, shared set of artifacts of practice, including cases, stories, theories, rules, frameworks, models, principles, tools, lessons learned, and of course, best practices. In this way, the full potential of the Kirkpatrick taxonomy can be realized.

In particular, it is suggested that these case studies include examination of three critical components of the evaluation system: (a) the roles evaluators play in the evaluation process and the competencies and expertise that they must bring to the process; (b) techniques and processes used to conduct the evaluations, especially in regard to strategies for time saving, cooperation, collaboration, and communication between training professionals and the business owners whose employees' behavior and operations' results are in question; and (c) the use of enabling technology, especially in regard to data access, collection, analysis, and reporting, and also including structure and definition of metrics and measurement classes.

Summary Reflections

These final paragraphs mark the end of a long journey for this researcher. This study began with an inquiry into the value of employee training and development as perceived by the enterprises that support such endeavors. That led to a detailed, and it is hoped, useful examination of Kirkpatrick's four levels, which is the training industry's standard for evaluating the effectiveness of training programs. This study now ends with an idea for how to make the important task of evaluation more widespread in use and more valuable in its application.

The general idea is very simple and well known, almost to the point of being a cliché: begin with the end in mind. Almost 50 years ago Donald Kirkpatrick proposed a taxonomy for evaluation that would give training professionals a way to get started. He suggested, indeed he insisted, that evaluators begin by measuring the reactions of the students to the training. He hoped that from this starting point, evaluators would go on to measure the greater complexities of learning, behavior, and results. But it hasn't quite worked out that way, as has been shown by these findings. Nonetheless, these findings also offer the possibility of a new approach, based on the success of those practitioners who have used Kirkpatrick's taxonomy. In the process, these practitioners have strengthened the taxonomy, making their evaluation efforts more useful and more valuable. So it is that perhaps the lesson learned from this study of Kirkpatrick's four levels is simply this: right steps, wrong order. Perhaps by encouraging training professionals first to understand an organization's desired results and working backwards from there, organizations will make investments in training programs with greater confidence and certainty that students will be satisfied, that they will learn, that they will perform to expectations, and that in the end, the organization will achieve the desired results.