

Kirkpatrick's Techniques for Evaluating Training Programs

Sevinj Iskandarova

James Madison University

September, 2015

Table of Contents

Introduction3

Part I – Summarize the article`s content3

 Kirkpatrick’s four level model4

 Kirkpatrick’s four level model definitions4

 Kirkpatrick’s four level model process and step-by-step explanations5

Part II – Why the articles are important for students in AHRD 640.....9

 Why is it important to enhance our understanding in M&E10

 How important is the article to the field of AHRD11

Part III –My personal thoughts on the articles11

Part IV14

 Kaufman and Keller (1994).....14

 Holton (1996)15

Part V - Conclusion15

References17

Introduction

The purposes of this paper are to write a summary of my understanding of Kirkpatrick's model. Secondly, I will summarize and comment on the additional articles as they critique Kirkpatrick's model. Finally, I will then reflect on each level of the model from my personal observations, incorporating my questions to the authors and readers of this paper seeking further explanations and deeper understandings.

Part I - Summarize the articles' Content

I believe that first of all we need to remember that when Donald Kirkpatrick published this article it was 1959. I am not sure when he started his initial thoughts or his initial work, or even when he crafted the idea in his head. But I believe this information is worth mentioning since his work is over 56 years old and is still being used today. That puts him right there with others whose works have stood the tests of time.

As training professionals, trainers are also considered lifelong learners. As lifelong learners, they often share a common space with managers, project stakeholders, subject matter experts (SMEs) and line managers; and are encouraged to learn from one another and demonstrate the new knowledge, skills and/or attitudes (KSAs), essential for the actual work setting. Learning from one another, while easier said than done, comes with its challenges when evaluating the effectiveness of training programs, mentoring programs, annual performance reviews and development activities. According to Donald Kirkpatrick and James Kirkpatrick, evaluating the effectiveness of a program is needed for "future program improvement, decisions to keep or drop a program and to justify the training department" (2006, p 96).

Kirkpatrick (1959a) has provided a four-level model of evaluation to assist training directors in evaluating programs within the organization, institutional and also governmental

areas. Kirkpatrick (1959a) proposes that the effectiveness of training increases organizational value and should be measured by the program’s return-on-investment.

Figure 1: Kirkpatrick’s (1959a) four step process to evaluation is visualized below:

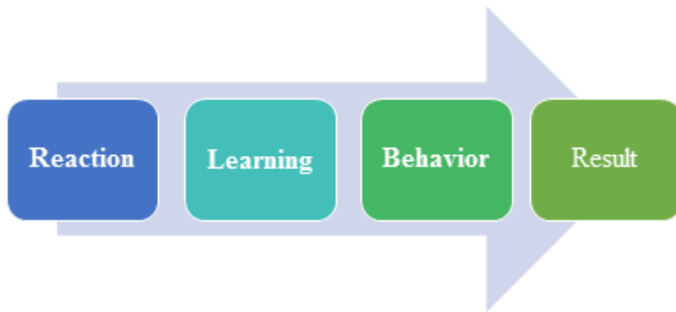
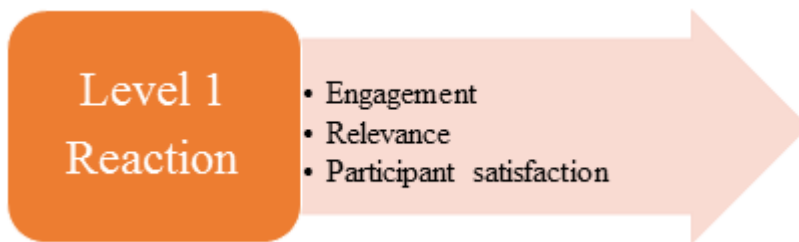


Figure 2: Kirkpatrick’s level model definitions

Level 1: Reaction	Participants react favorably to the training.
Level 2: Learning	Participants acquire intended knowledge, skills and attitudes.
Level 3: Behavior	Participants apply what they learned during training on the job.
Level 4: Results	Targeted outcomes occur.

Figure 3: Visualized Level 1 Reaction



The first level of Kirkpatrick model (1959a) identifies and describes the reaction process and measures participants’ reactions toward the training. Kirkpatrick states that “reaction is basically a measure of (internal) customer satisfaction and ensures that participants are motivated

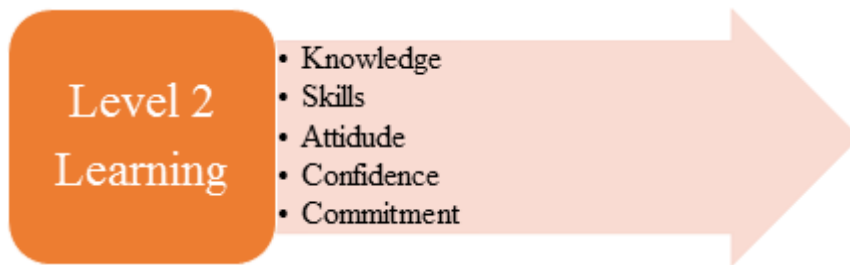
and interested in learning” (1959a, p 5). Reaction might also be described as measuring participants’ feeling and perception about the training.

Kirkpatrick has also explained that specific standards should be met when a training director evaluated trainees’ reaction. The standards are:

“Determine what you want to find out; use a written comment sheet with the items determined in the task above; design a form so that reactions can be quantified; obtain honest reactions by making the form anonymous; allow trainees to write additional comments not covered by the questions designed to be quantified (Kirkpatrick, 1959a p.4-5)”.

Regarding the evaluation of the first level (reaction), Kirkpatrick (1959a) suggested an evaluation techniques form (feedback form) to measure how well the trainer covered the training objectives during the training. The techniques form measures training based on a five skills check bar (excellent, very good, good, fair, and poor). This evaluation technique is very easy and quick in identifying participant’s reactions, and is usually done at the end of training. Another suggested evaluation technique is a comment sheets (verbal reaction). This technique is used during the last part of the training program by measuring participants’ reaction through verbal reaction, post-training surveys or questionnaires. The main part of this evaluation technique is that people have a positive impression of their training experience. This evaluation technique is not expensive nor is it time consuming (Kirkpatrick, 1959a).

Figure 4: Visualized Level 2 Learning

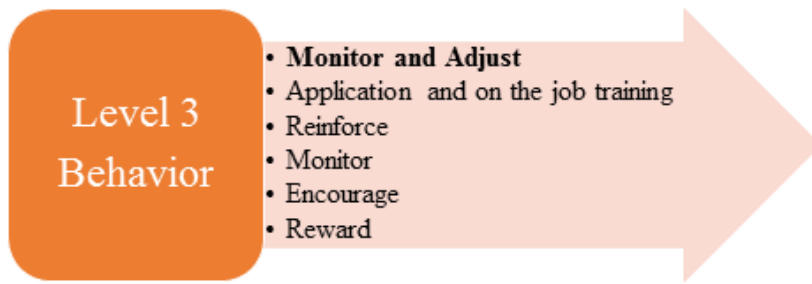


The second level of Kirkpatrick’s model (1959 b) measures participants improved knowledge, skills and attitudes as a result of attending training. The purpose of this level is to define learning principles, facts and techniques that were understood and absorbed by participants while still in the training (Kirkpatrick, 1959b). This level of evaluation provide information about trainers’ performance in increasing knowledge or changing attitudes, and is also based on before and after test results trainers might use to determine if the training objectives were met (Kirkpatrick, 1959b). Kirkpatrick has also explained that the specific standards should be met when a training director measures the learning. The standards are as follows:

“ measure the learning of each trainee so that quantitative results can be determined; use a before and after approach so that learning can be related to the program; as much as possible, the learning should be measured on an objective basis; where possible, use a control group (not receiving the training) to compare with the experimental group that receives the training; where possible, analyze the evaluation results statistically so that learning can be proven in terms of correlation or level of confidence” (Kirkpatrick, 1959b, p. 21).

The above guidelines are very important for level 2 evaluation, because these steps guide trainers to measure how much learning occurred (Kirkpatrick, 1959b).

Figure 5: Visualized Level 3 Behavior



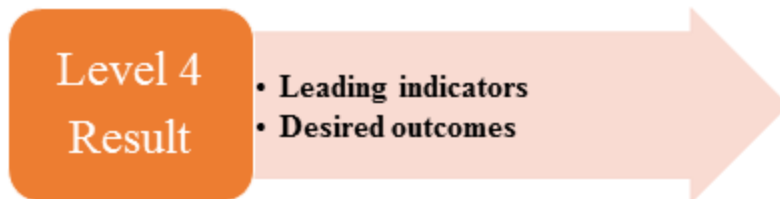
The third level of Kirkpatrick's model (1959c) measures the extent to which content learned in training was transferred to the job. Kirkpatrick mentioned that there is a difference in knowing principles and techniques and the ability to apply them in the actual work environment (1959c). Katz, (1956) in his article on *Human Relations Skills Can Be Sharpened*, stated that there are five requirements that participants must follow in order to change behaviors, which appears to be key at this level. The requirements are: participants must want to improve, recognize their own weaknesses, work in a permissive climate, they must have help from someone who is interested and skilled; and also, they must have an opportunity to try out new ideas. Kirkpatrick also noted:

“A systematic appraisal of on-the-job performances on a before-and-after basis; the appraisal of performance should be made by one or more of the following groups (the more the better): trainees; trainees' supervisors, subordinates, and peers and others familiar with trainees' on-the-job performance; a statistical analysis to compare before-and-after performance and to relate changes to the training; a post-training appraisal three month or more after training so that trainees have an opportunity to put into practice what they learned. Subsequent appraisals may add to the validity of the study; a control group should be used” (Kirkpatrick, 1959c, p.14).

After explaining the guidelines around behavioral changes, Kirkpatrick (1959c) described successful evaluation studies as cited by “Observed Changes Inquiry” study by Olav Sorensen

and Tarnopol (1959) approaches (Kirkpatrick, 1959c). The purpose of the description is to analyze practically that to “determine the effectiveness; attempts must be made to measure training in scientific and statistical terms” (Kirkpatrick, 1959c, p. 15).

Figure 6: Visualized Level 4 Result



The fourth level of Kirkpatrick model (1959c) is to evaluate when the discussed result was achieved. As he mentioned, this level is the most difficult to determine; simultaneously, it is the highest level of evaluation to achieving the desired results. The level typically measures the evaluation process, total quality management, economic trends, and year end promotions, leadership initiative and competitive environment (Kirkpatrick, 1959d).

In Likert’s (March/April, 1958) article, it is suggested that changes in productivity might be measured through pretest – posttest to determine whether the training program achieved its desired results or not (Kirkpatrick, 1959d).

In summary, Kirkpatrick mentioned how training program evaluation is important and beneficial for the real work environment; and he also discussed how his four-level model can be put effectively into practice. As I stated from the beginning at the time this model was developed there was nothing else in place. The model was good at the time and it meets the needs of many organizations. Only after approximately 30 years later researchers expressed concerns with model. Although there are many measurement techniques (Economic benefits or human good, Hamblin, 1974; Societal value, Kaufman and Keller, 1994; Return on investment

(ROI) Phillips, 1995; A six-level model, Brinkerhoff, 1987) attempting to “fix” Kirkpatrick’s model, there is skill evidence today showing the model is being used just as Kirkpatrick designed it to be used (Kirkpatrick, D.L. , 2009).

Part II – Why the articles are important for students in AHRD 640?

These articles are important for students in AHRD 640 because these articles enhance the students’ knowledge and understanding of program evaluation techniques. These articles are also important for HDR students for a number of additional reasons. First of all they provide a historical approach for the ideas for designing and building evaluation tools for HRD, first came into play. From what I can see the initial thought was out of a need to simply get a feel for how training was going. Secondly, it is quite impressive to see how those researchers coming along over 30 years later evaluate the work of those before them. Finally, to read how some of the researchers compliment the work of Kirkpatrick is quite refreshing.

As a student, I began to think in terms of building on the works of others through my own research, as I add to the body of literature. On the other hand, to see those who simply state that Kirkpatrick was wrong, is a bit disheartening. Respectfully, I don’t feel that wrong is the correct word. I think those critics might do well to think of the times that Kirkpatrick was writing in. Maybe look at how businesses and organizations were run during the 1950’s compared to later times. I also think these articles provide and encourage us as HRD students to think about the possibilities outside of the traditional evaluation model. There may not be just one model that will always fit every situation. Finally, as future trainers, Kirkpatrick’s suggestion and proposed model may be a useful tool for students to apply in their future work environment.

Why is it important to enhance our understanding in M&E?

It is important to enhance our understanding of Measurement and Evaluation because so often the best of the best researchers might get them confused misrepresenting the data. Once you have confused the two it becomes very difficult to determine what you are measuring and/or evaluating. I start with a definition of the terms Measurement and Evaluation taken from our text.

“Program Evaluation is the systematic collection of information about the activities, characteristics and results of programs to make judgments about the program, improve or further develop program effectiveness inform decisions about future programming, and /or increase understanding” (Patton, 2008, p.38).

“Performance Measurement is a system for assessing the performance of development interventions, partnerships or policy reforms relative to what was planned in terms of the achievement of outputs and outcomes” (United Nations Fund for Population Activities, 2004, p.8).

It is quite possible to do one without the other. For example, in Kirkpatrick’s model level one, you might measure whether participants liked the training and stop there. That would be measurement (two said yes or one said no). Or, you might take it a step further and make some type of evaluative judgments by comparing what participants said to a standard or some type of plan. By doing both measurement and evaluation you get a more complete picture of what the data are actually telling you.

Also, based on our course objectives an understanding of measurement and evaluation would help us demonstrate the use of various program evaluation and measurement models;

perform major data collection approaches to performance improvement and program evaluation; conduct initial data analysis for HRD program evaluation; and, plan, organize and prepare to conduct an evaluation and measurement project (AHRD 640 course syllabus, 2015).

How important is the article to the field of AHRD?

The article is very important to the field of AHRD. For one, it keeps the dialogue regarding measurement and evaluation in the field of AHRD going. Secondly, researchers are not starting from the beginning attempting to create an evaluation tool. For the purposes of this assignment it appears that Kirkpatrick's model laid a foundation for everyone else's work. Although some may say there are obvious flaws in his design, his contributions to the field was huge and his work will long be remembered. As students we might look forward to adding to this body of work if given the opportunity or, we might follow where the research continues to go. The good news is that these models are open to "evaluation". No one has a monopoly as to where the research might go.

Part III – My personal thoughts on the articles.

Kirkpatrick's four-level model is a straightforward practical system for evaluation processes. To apply Kirkpatrick's four-level model might take longer depending on desired results and the size of the organization; but it will definitely be a steady process to achieve the desired result. After reading this article, readers may easily understand: the four -level evaluation model, why each step is very crucial for program evaluation, how each steps should be applied and how to analyses data from different perspectives to achieve desired results. Moreover, Kirkpatrick's four-level model is also important in the training evaluation process for

simplifying the complex processes of training evaluation and improved organizational performance.

Level 1 Reaction

I agree that reaction is important to measure, but I also agree that what is being measured must be clearly articulated. As the author said, showmanship is often what's evaluated on content. I think the current study should be on the same questions using open ended responses. So often people simply go through the motion and respond to each questions without thinking through their response. I have even seen where prizes were given for completing the comment sheet. This strategy might influence the quality of the feedback, especially when names are used on the comment sheet. Sometimes the comment sheet provided more information than the presentation. And also sometimes questions on the comment sheets are designed to lead the participant to a certain response.

“Like” is relative. No two people will define the word in the same way. The researcher asking the question too, might have a different meaning for the word - What is it they “like”. These responses can apply to many aspects of the training. The reason for “like” the program may have nothing to actually do with the program goals. On the other hand, when you ask the question regarding “like” the question does open itself to qualitative a data analysis, allowing the researcher to potentially learn something that he/she may not have learned had he not gone a step further.

Dr. Steinmetz (1959a), commented that “people must like a training program to obtain the most benefit” (p.55). I am not sure that I agree totally with his statement. Although the training should be interesting and motivating those two are relative terms that might be further explored. I

believe that the content is the main focus and sometimes content does not lend itself to being entertaining. An example of this might be when training medical doctors on the most recent surgery technique, where procedures takes place over entertainment. Based on Malcolm Knowles adult learning participants are internally motivated by the content and the relevance the content is to their work/life.

I don't think Kirkpatrick (1959) intentionally left steps out as he was designing; I think he simply wanted to make sure that participants liked the training thereby deeming that they would be motivated to learn what was being presented. Also, if they liked the training they would go back to their jobs happy and change their behavior. The changed behavior would result in improved bottom line resultsfor the organization.

Level 2 Learning

Before and after approaches are wonderful for keeping everyone on the same page. If used to guide the training, it will definitely show participants how the intervention has helped them increase their content knowledge. This approach provides hard data for use in improving training and improved learning outcomes. Although the author states that a statisticion is needed, the trainer can make meaning of the data in a more quantitative manner by comparing the before and after responses or by simply sharing the before and after result.

Level 3 Behavior

Behavior level is also known as transfer to the job. What`s discussed in the articles happen so often in the workplace. People know what is expected often the correct things to do, but they often do what they have always done. On the other hand some people for one reason or another simple are not able to make the transfer from the training room to the job.

Level 4 Result

I totally agree that one way to measure effectiveness is in terms of the desired result, such as “reduced costs, higher quality, increased production and lower rates of employee turnover and absenteeism” (Kirkpatrick, 1959 d, p.59). On the other hand when one attempted to assess human behavior in the workplace time becomes a factor as well as whether a training actually took into consideration all of the factors of the actual work environment.

Part IV

Kaufman and Keller (1994)

There are also some critics who think there are major limitations with four-level model. Kaufman and Keller (1994) propose an extension for the four-level model by considering internal and external consequences of all interventions associated. Kaufman and Keller (1994) suggested adding one more level (as a fifth level) and named it “Kirkpatrick- plus” approach- “organizations would be encouraged to consider the societal consequences and payoffs of their actions” (p.377). Kaufman and Keller supported the researcher Senge’s (1990) approaches and they also mentioned that the reason for this addition is so that “every organization is a part of a large system with which it has mutually dependent relationships” (p.377). In addition, this extension would improve organization and human performance professionals’ performance in macro (the organizational level), micro (individual and/or small group) and mega (external clients and the society) levels.

Figure 7: Visualized Kaufman and Keller’s extended five levels of evaluation (1994) vs Kirkpatrick four level model (1959)

<p style="text-align: center;">Kirkpatrick’s Four Levels of Evaluation</p>	<p style="text-align: center;">Kaufman and Keller’s Five Levels for Evaluation of Interventions for Human Performance Improvement</p>
---	--

1. Reaction	1 a. Enabling 1 b. Reaction
2. Learning	2. Reaction
3. Behavior	3. Application
4. Result	4. Organizational Outputs
	5. Societal Outcomes

I do not agree with Kaufman and Keller's thought (1994), because to measure societal impact is a hard process. I do not think organizations will succeed in this step. Dr. Rees (2012) also mentioned that "if you think measuring organizational payoff is challenging, imagine how hard it is to measure societal impact" (p.3)

Holton (1996)

Holton (1996) also argued that Kirkpatrick's four-level model is incomplete. He complains that the four-level model needs to "develop further a theory of evaluation", because the four-level model doesn't fully identify all constructs. (Holton, 1994, p. 6). In regards to Kirkpatrick's model being incomplete, Holton (1996) claimed that four-level model is oversimplifies, the model doesn't cover essential elements, thus this noticeable absences affects "learning and transfer process such as trainee readiness and motivation, training design, reinforcement of training on the job" (Holton, 1996, p.7).

I would agree with Holton's argument (1996) that training reaction should be removed from evaluation models as a primary outcome of training. Because, measuring participant's reaction might too influence the quality of the feedback, especially when names are used on the comment sheets. And also sometimes questions on the comment sheets intentionally or unintentionally are designed to mislead the participant.

Part V- Conclusion

In conclusion, evaluating training programs are important in all organization. However, most organizations don't make the effectiveness of training a priority. Kirkpatrick (1959a) has provided a four-level model of to assist training directors in evaluating programs within the organization, institutional and also governmental areas. The four steps in Kirkpatrick's (1959a) process covers participants' reaction, participants' acquired knowledge, skills and attitude, if behavior changed when participants applied what they learned during training back on the job, and if the training reached the desired results. Kirkpatrick (1959a) also proposed that the effectiveness of training increases organizational value and should be measured by the program's return-on-investment. Whether one chooses to use Kirkpatrick's model or a revision of Kirkpatrick's model, this model is definitely a great place to start.

References:

AHRD 640 syllabus (2015). Retrieved from:

<https://canvas.jmu.edu/courses/1339473/assignments/syllabus>

Alternatives to Kirkpatrick: Kaufman's 5 Levels of Evaluation (2012). Retrieved from:

<https://instructionaldesignfusions.wordpress.com/2012/05/08/alternatives-to-kirkpatrick-kaufmans-5-levels-of-evaluation/#comments>

Hannum, Wallace. Training Myths: False Beliefs that Limit the Efficiency and Effectiveness of Training Solutions, Parts 1 and 2. *Performance Improvement*, Vol. 48, No. 2, 26-30 February 2009. Part 2: Vol. 48, No.6, July 2009

Holton, E. F. (1996). The flawed Four-level evaluation model. *Human Resource Development Quarterly*, 7, 5-21.

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

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

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

Kirkpatrick, D. (1960b). Techniques for evaluating training programs: Part 4—Results. *Journal of the American Society for Training and Development*, 14, 29-32.

Kaufman, R. & Keller, J. M. (1994). Levels of evaluation: Beyond Kirkpatrick. *Human Resource Development Quarterly*, 5, 371-380.

Kirkpatrick's Level 3: Improving the Evaluation of E-Learning (2005) Retrieved from:

<https://www.td.org/Publications/Newsletters/Learning-Circuits/Learning-Circuits-Archives/2005/05/Kirkpatrick's-Level-3-Improving-the-Evaluation-of-E-Learning>

Who's Afraid of Kirkpatrick's Four Levels? (2014) Retrieved from:

<https://www.td.org/Publications/Newsletters/Links/2014/02/Who-Is-Afraid-of-Kirkpatrick's-Four-Levels>

Kirkpatrick, D. L. & Kirkpatrick, J. D. (2006). *Evaluating training programs: The Four Levels* (3rd ed). San Francisco, CA: Berrett-Koehler

Katz, R. L. (1956). Human Relations Skills Can Be Sharpened. *Harvard Business Review*, 34(4),61-72.

Kirkpatrick, D.L. (2009). *Are you REALLY using the Four Levels?* Kirkpatrick Partners.

Retrieved from www.Kirkpatrickpartners.com

Knowles, M. (1973). *The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development*, Butterworth-Heinemann, NY

Russ-Eft, D., & Preskill, H. (2009). *Evaluation in organizations a systematic approach: a systematic approach to enhancing learning, performance, and change*. Basic Books. NY.