Field Epidemiology
In Action

Training Evaluation Theory

James Flint

Introduction

While evaluation theories, models, and methods are ubiquitous and continue to develop, many training programs are inconsistently, superficially, and poorly evaluated. Limited budgets, low organizational priority, and a lack of knowledge on how to properly conduct evaluations have been identified as key barriers. (1) Evaluation is fundamentally the process of determining the value, worth, and significance of a particular activity against set standards. In the training context, evaluation has been defined as the application of systematic methods to address questions about program operations and results to enhanced knowledge about underlying effectiveness and usefulness of the program. (2)

There are two main categories of training evaluation:

1. Summative: evaluations that are conducted after the program is complete to inform the decision about whether to continue, replicate, or scale up a program
2. Formative: evaluations that are occurring during the training for internal quality improvement and guiding program improvements

A large number of evaluation models reflect the diversity of the training programs being evaluated, the context in which they operate, and the specific focus of the evaluation. The following evaluation frameworks and models, prominent in training evaluation, all include aspects that focus on outcomes and/or impacts.

Kirkpatrick's Model

Kirkpatrick's Four Levels of Evaluation Framework
The best-known and most widely used theoretical framework for training evaluation originated with Donald Kirkpatrick. Kirkpatrick developed his four-step model in 1959 and it provides a pragmatic and straightforward approach for evaluating training programs. (3) While he has updated the evaluation framework several times over the years (in 1976, 1996, 1998, and 2006), the underlying model has remained the same. (4-7) The four levels of Kirkpatrick's evaluation model include:

1. **Reaction:** what are the feelings and reactions of the learners to the training? To what extent are the learners satisfied with the material taught and how it is taught?
2. **Learning:** to what extent did the learner master the educational content of the training? Are there signs of change in attitude, belief, or opinion?
3. **Behaviour:** to what extent has the learner demonstrated changed behaviour in the skill or knowledge areas covered by the course material? Is there evidence of the transfer of skills and knowledge to the workplace?
4. **Results:** to what degree has the learners' organization been positively impacted by the learners' application of the training, knowledge, and skills in the workplace?

The simplicity of the model has stood the test of time and it has become the de-facto framework for many evaluation frameworks. It remains the most popular evaluation model in the world today (8-11) and has been commonly used for evaluating medical and health-related training programs. (12)

It has, however, been criticised for not explicitly exploring the link between training and outcomes. (13) Kirkpatrick assumes a causal linear relationship between the four steps. Critics claim that it fails to adequately acknowledge the role of intervening and external variables, which may influence training outcomes. In response, several models emerged purporting to resolve some of these difficulties whilst retaining the core elements within Kirkpatrick.

Kirkpatrick Related Models

Many training evaluation models proposed over the last 50 years have used Kirkpatrick's four levels as their basis. This section summarises the most frequently cited models that have adapted or built upon the Kirkpatrick model.

CIRO Model: Warr, Bird, and Racklam – 1970

The CIRO (Context, Input, Reaction, Outcome) model includes an assessment of the context within which the training is conducted before assessing reactions and outcomes. (14) It also includes consideration and assessment of the inputs into the training program. The four stages of CIRO are:

1. Context – the operational situation that helps determine the training needs and objectives
2. Input – information about the possible training methods or techniques
3. Reaction – participant views on training program
4. Outcome – results of training in an immediate, intermediate and ultimate level

Hamblin's 'five-level' Model - 1974

One of the first to modify Kirkpatrick's model was Hamblin. (15) The first three levels in Hamblin's model resemble Kirkpatrick's model closely. The main difference is the fourth level, which Hamblin splits into organization and ultimate value. Hamblin suggests that the five levels of his model form a hierarchy. They are:

1. Level 1: Reaction - very similar to Kirkpatrick's model, where the trainers ask questions about the learners' reactions to the course immediately following the training.

2. Level 2: Learning - determining what the learners have learned in knowledge, skills, and attitude through the course; also very similar to Kirkpatrick's model.
3. Level 3: Job Behaviours - evaluating any change in job performance as a result of the learning in the course. Again, similar if not identical to Kirkpatrick's model where the trainer assesses the changes in the job performance of the students after the training
4. Level 4: Organization value - determining the effect of the training on the organization as a whole, such as a cost benefit analysis.
5. Level 5: Ultimate value - evaluating how the training has affected the "ultimate profitability and/or survival of the organization." This overall level is not included in the Kirkpatrick model.

The Organizational Elements model: Kaufman, Keller, and Watkins - 1995

Kaufman and Keller (1994) considered Kirkpatrick's model as too narrowly focused with limited application to practitioners interested in evaluating other types of development events. (16) They expanded Kirkpatrick's model to include societal contribution as an important evaluation criterion. They also included a needs assessment, the examination of the desired or expected results before commencing the evaluation, and a review of the availability and quality of resources used for the training. With the help of Watkins in 1995, the team developed their model with the following six levels.

1. Level 1: Input – an assessment of the role, usefulness, appropriateness, and contributions of the methods and resources used for the training.
2. Level 2: Process – similar to Kirkpatrick's reaction level including an analysis of whether the training intervention/program was appropriately implemented to achieve the stated objectives.
3. Level 3: Micro (acquisition) – similar to Kirkpatrick's learning level, examining individual and group mastery and competence.
4. Level 4 Micro (performance) – similar to Kirkpatrick's behaviour level, examining the utilization of skills and knowledge. The focus is on application rather than the transfer of skills and knowledge.
5. Leve 5: Macro – related to Kirkpatrick's results level, examining organizational contributions and payoffs.
6. Level 6 – Mega – an additional level that considers societal outcomes.

Phillip's ROI Evaluation Model: Phillips and Phillips - 1995

Phillips and Phillips suggested adding return on investment (ROI) to Kirkpatrick's four levels of evaluation. (17) While Kirkpatrick focused on results against stakeholder expectations (ROE), Phillips adds a cost-benefit analysis to determine the value of the training. This is very similar to Hamblin's 'organizational value'. This addition addressed the growing need for many training programs to justify their worth in the business setting. Return on investment focuses on monetary returns and while its immediate value can be seen within the for-profit sector, there may also be a role in the not-for-profit sector (including field epidemiology training programs). However, estimating monetary impacts of these training programs will likely be much more difficult.

Halton's Evaluation Research and Measurement Model - 1996

Holton, arguing that Kirkpatrick's four-level framework lacks a research foundation, proposed his model consisting of three primary outcome measures; learning, individual performance, and organizational results. (18) Holton suggested there was no correlation between reaction and results; therefore, did not include reaction in their model. Other evaluators have likewise rejected the link between reaction and results, arguing that positive reactions to training are not a predictor of learning. (19) Holton's model was later updated; however, was never widely used.

Indiana University Approach: Molenda, Pershing, and Reigheluth – 1996

Indiana University developed a six strata framework for evaluation; the strata were not designed to be a hierarchy of importance. (20) Their model builds on Kirkpatrick mainly by adding a level (stratum) before and after Kirkpatrick's four levels. The first stratum, called activity accounting, examines the training volume and level per participant. The sixth stratum examines the impact on society, along the same lines as Hamblin's 'ultimate value' and Kaufman's 'societal impact' levels.

KPMT model: Kearns and Miller - 1997

The KPMT model places emphasis on the identification of business needs, rather than training needs, when designing the training cycle. (21) Kearns and Miller differ from other models in their belief that return on investment can only be looked at in hard terms. They argue that training can only bring added value if the organization is not performing effectively, or there is a market opportunity to be exploited. They state that if a business objective cannot be cited as a basis for designing training and development, then no training and development should be offered. They include an initial phase where the business needs are examined and solutions designed with buy-in from management and stakeholders. The evaluation levels are very similar to Kirkpatrick's:

1. Reaction to the training and development
2. Learning
3. Transfer to the workplace/behaviour
4. Bottom line added value, measured to the base level measurements taken

Kraiger's Decision Based Evaluation - 2002

Krieger's model emphasized three target areas for evaluation of training. (22) Namely, training content and design (including delivery), changes in learners (including affective, cognitive, and behavioural) and organizational payoffs. The model encourages tailoring evaluation measures to the needs of the intended audience.

Integrated Model of Training Evaluation and Effectiveness (IMTEE) - 2004

Building on the work of Kirkpatrick, a new model based on a 20-year analytical review of research studies in program evaluation was developed by Holton and Kraiger. (19) The model assesses program effectiveness and program evaluation together. Evaluation measures relating to post-training attitudes were learning, training performance, and transfer performance.

Scriven's Model - 2010

Scriven added several levels to Kirkpatrick to create a 12-point evaluation checklist. (1, 23) His evaluation checklist points are:

1. Need: evidence that the proposed training is the best answer to a real problem at this time and for this group.
2. Design: the training design is appropriate for the demonstrated need, the target group's background and the resources available at the planned delivery site.
3. Recruitment: how well does the recruitment of trainees from the program include those who need it and exclude those who don't, acknowledging that recruiting less than ideal candidates has opportunity costs and may impede the efficiency of the training for others.
4. Delivery: evidence that the training was attended, supported and presented as proposed.
5. Reaction: as per Kirkpatrick.
6. Learning: as per Kirkpatrick.
7. Retention: how well did the participants retain their learning, knowledge, skills, attitudes or values.
8. Application: similar Kirkpatrick's level 3; how well have trainees appropriately used, and continue to appropriately use, what they have learned in their work context. Assessment requires one or more of the following activities – observation of work performance, an examination of work products, interview of the supervisor and interview of co-workers.
9. Extension: can the training model be replicated for other training purposes or in other contexts.
10. Value: the qualitative value of the impact of the training, both intended and unintended.
11. Alternatives: comparison of the impact of the training as compared to the (measured or estimated) impact of known alternative approaches to meeting the same needs as the training addresses.
12. Return on Investment: the overall return on investment in terms of economic environmental, ethical, extension (use in other contexts) and exclusiveness (comparative value).

Scriven also suggested that an external review of the evaluation itself is a valuable inclusion for any evaluation framework. Having an evaluator review the methods, Scriven suggests, is one of the best investments one can make in improving the evaluation of training.

New World Kirkpatrick Model - 2016

After the death of Kirkpatrick, his son and daughter-in-law published the 'New World Kirkpatrick Model' which modernized the four levels to maximize their effectiveness. (8) This revised Kirkpatrick model addressed some of the limitations highlighted by critics as it attempted to address the complexities of today's learning environment.

One of the criticisms of the earlier Kirkpatrick model was that most investigators in training evaluation stopped at level 2 (learning). (12, 24) The New World Kirkpatrick Model suggested evaluating the levels in reverse, concurrently or non-sequentially. There is a recognition that not all training needs to cover all levels and scarce evaluation resources should focus on the

critical evaluation objectives, not the order in which they should be conducted. Critics of the original Kirkpatrick model also argued that evaluation of levels 3 and 4 was too difficult. (24) The New World model suggests that the challenges of levels 3 and 4 evaluation often arise because evaluation is an afterthought, or the outcomes are not well aligned to the program or its key stakeholders. They emphasize the importance of identifying level 3 and 4 outcomes in collaboration with the program stakeholders during the planning, and jointly developing indicators for evaluation. They also include a broader description of what to evaluate at level 3 and 4, including the processes that facilitate the application of learned knowledge and skills. This revised model emphasizes the importance of on-the-job learning and includes this as a key element in the evaluation framework.

Another criticism of the earlier Kirkpatrick model was that it inadequately addressed confounding variables that may influence the given outcomes, such as learner motivation, variable entry levels of knowledge and skills. (13, 18, 25) The New World model acknowledges the impact of learner characteristics and organizational resources on the success or failure of programs. The original model has been expanded to include an assessment of the 'required drivers' (the processes to reinforce, monitor, encourage and reward learners for applying learned knowledge and skills), 'on-the-job learning' (inclusion and acknowledgment of the learning which occurs outside the classroom and the sense of responsibility and motivation to improve or change their practice), 'confidence and commitment' (learners perception that they intended to and will be able to apply learned knowledge and skills), and 'engagement and relevance' (the degree to which learners engage in, contribute to and have opportunities to use learning). The New World model focuses on evaluating chains of evidence rather than causal chains and encourages evaluators to identify and assess all factors, including those outside the training program, that influence outcome. The New World model does not address unintended outcomes, which remains a limitation of this model, especially in the field of medical and health education.

World Health Organization Training Evaluation Framework - 2010

The World Health Organization published their training evaluation framework in 2010. (26) It is based on Kirkpatrick and Phillips evaluation models and includes five levels of evaluation, as outlined in Table 2. The definition of level five is similar to the 'societal' element in Kaufman's Model and is pushing towards assessing the ultimate impact of a health-related training program; improved health outcomes in a population. While the WHO model includes tools and templated for levels 1-4, there is little information describing how to conduct the level 5 evaluation with no approaches, tools, or templates provided.

Table 2. WHO Training Evaluation Framework, 2010

LEVEL	FOCUS	AREA OF IMPACT	EVALUATION QUESTION
1	Reaction	The reaction of trainees	Are the trainees satisfied?
2	Learning	Increase in trainees knowledge and skills	What have the trainees learned?
3	Behaviour	How trainees apply their knowledge and skills	What do the trainees do differently in the workplace?
4	Results	How the training affects the trainees' broader area of work	What is the effect on the output of the trainee's team or department?
5	Return on investment	Wider impact on the achievement of public health programme objectives	How has the training contributed to achieving public health objectives?

Alternate Models

There are a number of evaluation models that were formulated without reference to Kirkpatrick. The following section highlights some of the most popular ones used for training evaluation.

CIPP Evaluation Model: Stufflebeam et al - 1971

The CIPP evaluation model is structured around four parameters: Context, Input, Process, and Product. (27) Developed by Daniel Stuffelbeam and colleagues in the 1960s, its focus was on improving the relevance of evaluation for the decision-maker. Context evaluation determines the alignment of the training with the expressed organizational need and formulates objectives in light of those needs. Input evaluation helps focus on the designing of the training by examining capability, resources, and different training strategies. Process evaluation refers to the systematic monitoring of the training program and product measures the attainment of the training against the objectives. Product evaluation helps to judge and react to the program's outputs and outcomes. The model examines both the intended and unintended consequences of improvement efforts. Stufflebeam argues that process evaluation is essential to provide the basis for interpreting the reason for the outcomes.

The Learning Outcomes Approach: Kraiger et al - 1993, 2002

This model emphasizes training outcomes and suggests the need to distinguish between cognitive, skill-based, and affective outcomes when designing the evaluation. (22, 28) Distinguishing the outcomes can be done by viewing the instructional objectives through different lenses, with the different perspectives looking at the goals the training, the process strategies and the performance criteria. Kraiger questioned whether Kirkpatrick differentiates between skills and facts, since the model measures them with the same assessment tools. Kraiger argues that knowing a fact does not always mean that the person has the skills to demonstrate it. Therefore, different types of assessments are required to measure different types of knowledge.

Responsive Evaluation Model: Pulley – 1994

Responsive evaluation focuses on tailoring the evaluation and evaluation results to the needs of the decision makers. (29) Pulley argues that the objective of any evaluation should be to provide evidence so key decision makers can determine what they want to know about the program. The approach uses both quantitative and qualitative data and relies on stories and anecdotes to 'give life to the numbers'. The engagement and involvement of the decision makers throughout the entire evaluation process is a key element of the responsive evaluation model. Seeking their input to custom design an evaluation approach, rather than using a pre-existing framework, is central to the model.

Anderson Model of Learning Evaluation - 2005

This model was first published by the Chartered Institute of Personnel and Development in 2006 and differs from other training evaluation models in its focus on aligning an organization's training program(s) with its strategic priorities. (30) It is less concerned with specific learner benefits or the evaluation of individual training activities. This model uses a range of methods to assess and evaluate the contribution of learning to the organization's strategic priorities.

References

1. Scriven M. Types of evaluation and types of evaluators. Evaluation Practice. 1996;17(2):151-61.
2. Newcomer K.E. HHP, Wholey, J.S. Handbook of practical program evaluation. 4th Edition ed: John Wiley & Sons; 2015.
3. Kirkpatrick DL. Techniques for Evaluation Training Programs. Journal of the American Society of Training Directors. 1959;13:21-6.
4. Kirkpatrick DL. Evaluation of Training. In: RL C, editor. Training and development handbook: a guide to human resource development. New York: McGraw Hill; 1976. p. 301-19.
5. Kirkpatrick DL. Evaluating traiing programs: the four levels. 1st ed. San Francisco: Berrett-Koehler; 1996.
6. Kirkpatrick DL. Evaluating training programs. 2nd ed. San Francisco: Berrett-Koehler; 1998.
7. Kirkpatrick DL KJ. Evaluating training programs: the four levels. 3rd ed. San Francisco: Berrett-Koehler; 2006.
8. Kirkpatrick JD KW. Kirkpatricks four levels of training evaluation. Alexandria: ATD Press; 2016.
9. Sugrue B RRJ. State of the Industry: ASTD's Annual Review of trends in Workplace Learning and Performance. Alexandria, VA: American Society of Training and Development; 2005.
10. Aguinis H, Kraiger K. Benefits of training and development for individuals and teams, organizations and society. Annual Review of Psychology. 2009;60:451-74.

11. Twitchell S HEF, Trott J.W. Technical Training Evaluation Practices in the United States. Perform Improvement Quarterly. 2000;13(1):84-109.

12. Moreau KA. Has the new Kirkpatrick generation built a better hammer for our evaluation toolbox? Medical Teacher. 2017;39(9):999-1001.

13. Frye AW, Hemmer PA. Program evaluation models and related theories: AMEE guide no. 67. Medical Teacher. 2012;34(5):e288-99.

14. Warr P.B. BM, Rackham N. The Evaluation of Management Training Gower; 1970.

15. Hamblin AC. Evaluation and Control of Training. United Kingdom: McGraw Hill; 1974.

16. Kaufman R KJ, Watkins, R. . What works and what doesn't: Evaluaton beyond Kirkpatrick. Performance and Instruction. 1995;35(2):8-12.

17. Phillips J. J. P, P. . Measuring return on invesgtment in leadership development.

The handbook of leadership development evaluation2007.

18. Holton III EF. The flawed four-level evaluation model. Human Resource Development Quarterly. 1996;7(1):5-21.

19. Alvarez K, Salas E, Garofano CM. An integrated model of training evaluation and effectivness. Human Resource Development Review. 2004;3(4):385-416.

20. Molenda M PJ, Reigeluth C. Designing instructional systems. In: R. C, editor. The ASTD Training and Development Handbook: a Guide to Human Resource Development. 4th ed. New York: McGraw Hill; 1996.

21. Kearns P MT. Measuring the impact of training and development on the bottom line. London: Pitman Publishing; 1997.

22. Kraiger K. Creating, implementing, and managing effective traiing and development: State of the art lessons for practice: Jossey-Bass; 2002.

23. Scriven M. The evaluation of training: A checklist approach.; 2010.

24. Yardley S, Dornan T. Kirkpatrick's levels and education 'evidence'. Medical Education. 2012;46(1):97-106.

25. Bates R. A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. Evaluation and Program Planning. 2004;27(3):341-7.

26. Organization WH. WHO training evaluation guide. Geneva: WHO 2010.

27. Stufflebeam DL. CIPP Evaluation Model Checklist, A tool for applying the CIPP Model to assess long-term enterprises: Evaluation Checklists Project 2007 [cited 2020 Jan 7]. Available from: https://wmich.edu/sites/default/files/attachments/u350/2014/cippchecklist_mar07.pdf.

28. Kraiger K, Ford JK, Salas E. Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. Journal of Applied Psychology. 1993;78(2):311-28.

29. L. PM. Navigating the evaluation rapids. Training and Development. 1994;49(8):19-24.

30. V. A. The Value of Learning: A New Model of Value and Evaluation. London, England: Chartered Institute of Personnel and Development (Great Britain) (CIPD); University of Portsmouth Business School; 2007.