Advances in Economics and Business Management (AEBM)
p-ISSN: 2394-1545; e-ISSN: 2394-1553; Volume 2, Issue 11; July-September, 2015 pp. 1086-1090

http://www.krishisanskriti.org/Publication.html

Examination of Models of Evaluation Revalidation of Evaluation Models

Rama Reddy Goluguri

Independent Researcher Hyderabad, India
E-mail: reddy@ausi.com

Abstract—*In a dynamic ever-changing corporate environment, it can be impossible to rely on a single evaluation theory or model to measure specific returns realized following a substantial investment on a leadership learning and development program. In order to bridge gaps in the context of evaluation, it is required to revalidate existing theories and models. Moreover, developing a sandwich model may help obtain synergies through the value of training intervention in social and financial terms. It is needless to say that any evaluation without integrating it into the training design process would not yield the intended results.*

Index Terms*: Leadership, Learning & Development, Evaluation Models & Theory, Critical Evaluation, Revalidation. (key words)*

1. INTRODUCTION

Evaluation in the context of leadership, learning and development is an analytical process as it involves systematic collection of data to determine success or failure of leadership development programs in terms of quality, effectiveness or value created within the organization (Goldstein, 1986; Hannum and Martineau, 2008). In practice, evaluation occurs when specific measures of outcome are conceptually linked to intend learning objectives (Kraiger et al., 1993). Therefore, evaluation process should encompass the total value of a training intervention in social and financial terms (Talbot, 1992).

There are two types of evaluations: summative and formative. Summative evaluation is about proof of application of the learning and formative evaluation is about improvement as a result of learning (Michalski and Cousins, 2001).

In other words, formative evaluation is indicative of what has been learned, and summative evaluation is about whether or not it has been applied appropriately in the right context (Smith, 1986); (Alvarez, Salas, & Garofano, 2004).

According to Smith, there are four main purposes of evaluation: research, pragmatic, political and organizational learning (see Appendix 1). Research purpose is used to shed light upon the processes of learning and development in organizations. Furthermore, a pragmatic purpose provides feedback for learning and development practitioners and program sponsors. Moving forward, a professional-political purpose is specific to the organization, occupation and of course, the society as a whole. Organization's learning assumes the role of practitioners as change agents by engaging them in the processes of reflection and dialogue (Smith, 2006).

The overall aim of evaluation process is to influence corporate decisions about the need for the learning and development programs in the future, the need for modifications to the programs, and the need to provide cost/benefit data about the programs (Goldwasser, 2001).

This paper is intended to examine the effectiveness of existing evaluation theories and models with respect to organizations broader Human Resources Development Strategy. Thus, a working definition for evaluation that combines all the above mentioned elements has been adopted to proceed with further discussion in the following sections that includes systematic collection of descriptive and judgmental information necessary to make effective decisions relating to the selection, adoption, value and modification of various instructional activities (Goldstein, 1993).

2. CRITIQUE OF EXISTING THEORIES AND MODELS OF EVALUATION

This section will critically assess the validity and expose the limitations of existing theories and models of evaluation in relation to leadership learning and development. Many theories and models of evaluation have been developed since the 1960s. Within these some make a distinction between validation and evaluation. Validation is measuring whether the training has delivered what it promised (Critten, 1991). The earliest theories were based upon scientific principles of observation and measurement and sought to identify cause and effect.

In particular organizations wanted to establish the financial value of the training by examining costs and benefits (Bramley, 1991). This hard accounting approach was soon challenged by a more pragmatic, soft approach which sought to measure the value of training by identifying the indirect and immeasurable effects such as on morale and group dynamics

which earlier theories thought that could not be quantified (Fielder and Pearson, 1979; Bramley, 1991a).

A further advancement came with the concept of the training cycle that involves identifying training needs, or the training gap, choosing the appropriate training intervention, implementing it and evaluating the results of the training against the original training needs. This also came in for criticism because of its internal focus on both the training process and its evaluation and in particular the passive role of participants.

Subsequently, the Kirkpatrick framework for evaluation of training (Kirkpatrick, 1967), also known as the four-level (reactions, learning, behavior, results) evaluation model (see Appendix 2), is acknowledged by many practitioners as the standard framework in the field, but not the model.

The reason is on one hand it remained as framework because it does not attempt to explain the cause-and-effect relationship between different elements of the framework (Alliger & Janak, 1989). On the other hand the four levels of evaluation are not well grounded in research but is a prescriptive model (Holton and Naquin, 2005).

As far as first two levels are concerned Dixon demonstrated that there is little correlation between reactions and learning (Dixon, 1990); (Alliger, 1997). Moreover, it represents a trainer's notion of what constitutes effective evaluation not a business manager's (Twitchell et al, 2000). As with all strategist frameworks there is an assumption of linearity that the first stage causes the second and so on. Thus evaluation starts after the training has been done or during the training and not before (Kearns, 2005). No evaluation model can be validated without measuring and accounting for the effects intervening variables. Therefore, it can be argued that the four-level system of training evaluation-reaction, learning, behavior, results is a taxonomy or framework of out-comes and is incomplete to be considered as an evaluation model.

Furthermore, Holton argues the Kirkpatrick model is fundamentally flawed and incomplete because it takes no account of the many contextual factors that may affect the transfer of learning, including the ability and motivation of the trainee, the relevance of the training to the needs of the trainee and the organization, and the receptiveness of the workplace organization to the transfer of the learning (Holton, 1996). Hence, it does not meet the criteria for a theory or a model.

Other evidence to support earlier arguments are to be a number of subsequent inclusions to the Kirkpatrick framework as a measure to bridge the gaps caused by missing elements, including adding a fifth level to reflect trainings ultimate value in terms of organizations success criteria, such as economic benefits or human good (Hamblin, 1974a) and societal value (Kaufman and Keller, 1994), or to focus more specifically on return on investment (Philips, 1995). Further, Brinkerhoff proposed a six-level model that in essence added two formative evaluation stages as precursors to Kirkpatrick's four levels (Brinkerhoff, 1987).

All subsequent additions to Kirkpatrick framework can be best labeled as taxonomies, which are simply classification schemes (Bobko and Russell, 1991). According to Bobko and Russell, taxonomies are the link between the initial stages and final confirmatory stages of theory. Although the Kirkpatrick model is elegant in its simplicity and has contributed greatly to human resources development, the lack of research to develop a theory of evaluation is still a glaring shortcoming for the field.

By contrast to Kirkpatrick framework, CIRO frame work focuses on measurements taken before and after the training has been carried out. The key strength of this model is that it takes into account the resources of the organization. Tennant critiques CIRO model by highlighting that it does not take behaviours into account (Tennant et al, 2002). It can be argued that this model is suitable for managerial focused training programs rather than those that are less specialized and perhaps aimed at people working at lower levels in the organization.

Smith, in contrary, argues against causal assumptions of Kirkpatrick and CIRO frameworks. Further, he puts forward the framework: context, administration, inputs, process and outcomes (CAIPO) (Smith, 1986) as an alternative to demonstrate overlooked cause-and-effect chain in the former mentioned frameworks, especially with regard to ultimate level evaluations (Easterby-Smith, 1994). It can be argued that this framework strongly challenged the causal assumptions in the predecessors' frameworks. Despite of this frameworks additional effort toward bridging the relational gaps in predecessor's frameworks this also ended up as an augment to previous frameworks rather than a model.

In conclusion, there is a need for a unifying model for evaluation theory, research, and practice that will account for the collaborative nature and complexities involved in the evaluation of training. Furthermore, as a result, according to CIPD annual survey data, many training practitioners found that 'serious' evaluation was too time-consuming (CIPD, 2007). Interestingly, none of the available frameworks or so called models for training evaluation seems to be accountable for these gaps of evaluation.

It is evident from the above facts that existing frameworks fall short in comprehensiveness and it can be argued that they fail to provide tools that guide organizations in their evaluation systems and procedures. Not surprisingly, organizations and researchers are still in search of better solutions to help eliminate problems with respect to developing consistent evaluation models that last long.

3. CRITICAL EXAMINATION OF EFFECTIVENESS OF DOMINANT EVALUATION MODELS

The purpose of this section is to critically examine how dominant evaluation models can serve to improve the effectiveness of leadership learning and development within the context of organizations Human Resource Development (HRD) strategy. Moreover, increasing global competition has led to intense pressure on HRD to demonstrate that learning and development interventions contribute directly to the organizations bottom-line. Therefore, measuring training effectiveness became an important strategy of the organization.

There are some criteria for measuring the effectiveness of training: direct cost, indirect cost, efficiency, performance to schedule, reactions, learning, behavior change, performance change (Shepherd, 1999). Thus, critical examination offers a rich source of ideas about how to evaluate at different levels and amongst different stakeholders and what are the difficulties in doing so to improve the overall effectiveness of leadership learning and development.

Among the dominant evaluation frameworks, the first one to examine would be Kirkpatrick four levels of evaluation because of its wide acceptance of practitioners (Stone and Watson, 1999). This framework promoted the awareness of the importance of thinking about and assessing training in business terms (Wang, 2003). His four levels of evaluation are: reaction – a measure of satisfaction, learning – a measure of learning, behavior – a measure of behavior change and results - a measure of results (Phillips, 1997); (Kirkpatrick, 2009). These four levels cover much of the criteria for measuring success as mentioned earlier.

Reactions evaluation assists an organization to improve the effectiveness of leadership learning by assessing participant's reactions to a course's instructor, setting, materials, and learning activities. It can be argued that this level of training evaluation contribute to the overall effectiveness because it involves gaining direct feedback from the participants (Kirkpatrick partners, 2011a).

Learning evaluation involves determining the extent to which learning has occurred. Learning assessments include performance testing, simulations, case studies, plays, and exercises. It can be argued that a variety of techniques that are used for determining if the learning objectives have been met contribute to help improve the overall effectiveness of leadership learning (Kirkpatrick partners, 2011b).

Behavior evaluation attempts to determine the extent to which new skills and knowledge have been applied on the job. It can be argued that by not only ensuring satisfaction ratings are good and the learning objectives are met but also determining the extent of transfer of knowledge into behavior contributes to improve the overall effectiveness of leadership learning and development (Kirkpatrick partners, 2011c).

Results evaluation involves measuring system wide or organizational impact of leadership learning. It can be argued that by providing statistics relevant to performance of business measures such as improved output, improved quality, decreased costs, increased sales and less time consumption after the training help contribute toward improving the overall effectiveness of leadership learning activity (Kirkpatrick partners, 2011d).

Return on Investment (ROI) also known as fifth level to the Kirkpatrick framework is a measure of the monetary benefits obtained by an organization over a specified time period in return for a given investment in a training program. In other words, ROI is the extent to which the outputs of training exceed the inputs. It can be argued that ROI can be used both to justify a planned investment and to evaluate the extent to which desired return was achieved in order to improve the effectiveness of leadership learning. By calculating ROI on the courses where it is possible, it is more appropriate to be trusted on the ones that cannot evaluate at four levels (Parry, 1996).

CIRO was built upon Kirkpatrick' model identifying seven levels of which included context, input, reactions, outcomes (CIRO) but also three levels of outcomes which referred to as ultimate, intermediate and immediate (Warr, Bird and Rackham, 1970). Each level is concerned with acquisition of information in order to make effective decisions that in turn help improve the overall effectiveness of leadership learning and development activity. Firstly, context evaluation focuses on factors such as the correct identification of training needs and the setting of objectives in relation to organization culture, structure, support and climate. Secondly, input evaluation that is primarily related to design and delivery stage of the learning cycle focuses on possible resources and choice between alternative learning and development activities.

Thirdly, reaction evaluation looks at gaining and using information about the quality of trainees' experiences and immediate or delayed reactions to the process of learning and development. Finally, outcome evaluation focuses on the achievements gained as a result of learning and development in terms of job behavior and job performance. Subsequently, to measure the learning effectiveness results are assessed at three levels as explained below.

Immediate evaluation attempts to measure changes in knowledge, skills or attitude before a trainee returns to the job. Intermediate evaluation refers to the impact of training on job performance and how learning is transferred back into the workplace. Finally, ultimate evaluation improves the overall effectiveness of leadership learning by assessing the impact of learning interventions on departmental or organizational performance in terms of overall results (Hamblin, 1974b).

Smith introduced CAPIO framework as an alternative to Kirkpatrick and Warr et al (CIRO) frameworks by including and demonstrating cause-and-effect chain with regard to

ultimate level evaluations to help improve the overall effectiveness of leadership learning and development. His alternative framework includes context, administration, inputs, process and outcomes (CAPIO) (Smith, 1986).

Context evaluation focuses on factors outside and beyond the training program, for example, the level of support for learners at the workplace. Administration evaluation is concerned with the mechanisms of nomination, selection and briefing before any training starts, and any follow-up activities e.g. debriefing by the line manager or post-course evaluation. Evaluation of inputs examines the content and methods of training. Process evaluations focus on what actually happens during a training activity and how the participants experience it.

Finally, outcome evaluation is concerned with establishing the outputs or outcomes of employee training and development. Thus the frameworks main focus is on individuals and changes in their knowledge, skills, attitudes and behavior, individual performance, organization performance or on shifts in organization culture and climate. Methods used in applying the CAIPO framework are somewhat similar to those used in earlier frameworks. However, this framework provides a series of choices as explained above for evaluation to help improve the overall effectiveness of leadership, learning and development program.

4. CONCLUSION

In conclusion, evaluation models that do not return the knowledge and information deemed most useful for key clients and stakeholders are unlikely to be adopted for practise (Patton, 1997). In future, research needs to focus on the barriers to developing effective evaluation models by understanding how leadership learning is being evaluated and integrated with the training design, how the collaborative process of evaluation is being managed and how they may be assisted.

However, by and large, existing evaluation frameworks help managers, employees and human resource professionals to make informed decisions toward improving the effectiveness of learning interventions. But, the choice of selecting evaluation frameworks depends on the maturity and environmental constrains of the organization.

REFERENCES

[1] Approaches to Evaluation of Training: Theory & Practice available at http://www.ifets.info/journals/5_2/eseryel.html accessed on Nov 28, 2012

[2] Alvarez, K., Salas E., & Garofano, C. M. (2004) An integrated model of training evaluation and effectiveness, Human Resource Development Review, p385-416.

[3] Alliger, G. M., & Janak, E. A. (1989) Kirkpatrick's levels of criteria: Thirty years later, Personnel Psychology, p331-341

[4] Alliger, G.M., Tannenbaum, S.I., Bennett, W., Traver, H., and Shotland, A. (1997) A meta-analysis of relations among training criteria, personnel psychology, p341-358.

[5] Brinkeroff, R. O. (1987) Achieving results from training, San Francisco, Jossey-Bass

[6] Bobko and Russell (1991) A review of the role of taxonomies in human resources management, human resources management review, p293-316

[7] Bramley, P. (1991); (1991a) Evaluating Training Effectiveness, McGraw Hill.

[8] CIPD (2007) Survey reports available at: hhtp://www.cipd.co.uk/hr-resorces/survey-reports/accessed on November 30, 2012.

[9] Critten, P. (1993) Investing in People: Towards Corporate Capability, Oxford: Butterworth Heinemann.

[10] Dixon, N. M. (1990) The relationship between trainee responses on participation reaction forms and past test scores, Human Resources Development quarterly, p129-137.

[11] Easterby-Smith, M., (1994) Evaluating Management Development, Training and Education, 2nd Edition, Gower.

[12] Fielder and Pearson (1979) An Approach to Costing Training: Feasibility Study, London: Peat, Marwick and Mitchell.

[13] Goldstein, I. (1986) Training in organizations: Needs assessment, development and evaluation, B rooks/ Cole, Monterey, CA.

[14] Goldstein, L. (1993) Learning and Development for Managers, Sadler-Smith (2006), Blackwell Publishing, p375

[15] Goldwasser, D. (2001) Beyond ROI. Training, p82

[16] Hannum, K. and Martineau, J (2008) Evaluating the impact of leadership development, Pfeiffer, San Francisco.

[17] Hamblin, A. C. (1974a); (1974b) Evaluation and control of training, New York, McGraw-Hill

[18] Holton, E. (1996), The Flawed Four- Level Evaluation Model, Human Resource Development Quarterly, Vol.7, p5-21.

[19] Holton, E. and Naquin, S. (2005) A critical analysis of HRD evaluation models from a decision-making perspective, Human Resources Development quarterly, p257-280.

[20] Hamblin (1974) the Evaluation and Control of Training, London: McGraw Hill.

[21] Kraiger, K., Ford, J. K. and Salas, E. (1993) Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation" , Journal of Applied Psychology, vol. 78, no. 2, pp. 311-328.

[22] Kaufman, R. (1994) Levels of evaluation beyond Kirkpatrick, human resource development quarterly, p371-380

[23] Kearns, P. (2005) Evaluating the ROI from Learning, London: CIPD.

[24] Kirkpatrick, D. (1967) Evaluation of Training in Craig, R (Ed) Training and Development Handbook, Mc Graw Hill

[25] Kirkpatrick, J., and W. Kirkpatrick (2009) The Kirkpatrick Four Levels: A Fresh Look After 50 Years, 1959-2009, Available at:

[26] http://www.managesmarter.com/managesmarter/images/pdfs/trg_20090417_kirkpatrickwhitepaper.pdf, Accessed on Nov 2, 2012

[27] Kirkpatrick partners (2011a); (2011b); (2011c); (2011d) The Kirkpatrick Philosophy, Available at: http://www.kirkpatrickpartners.com/OurPhilosophy/tabid/66/Default.aspx Accessed on Nov, 4, 2012

[28] Michalski and Cousins (2001) multiple perspectives on training evaluation: probing stakeholder perceptions in a global network

development firm, American Journal of Evaluation, vol. 22, no. 1, and p37-53.

[29] PARRY, S. (1996) Measuring Training's ROI, Training and Development, p72.

[30] PHILLIPS, J (1997) Handbook of Training, Evaluation and Measurement Methods, Houston Gulf, p39.

[31] Patton, M.Q. (1997) Utilization-focused evaluation (3 rd ed.), Thousand Oaks, CA, Sage.

[32] Shepherd, C. (1999) Assessing the ROI of training available at: http://www.fastrack-consulting.com.uk, accessed on Nov 8, 2012

[33] Stone, J. and Watson, V. (1999) Evaluation of Training, Available at: http://www.ispi-atlanta.org, accessed on Dec 2, 2012

[34] Sadler-Smith, E. (2006). Learning and Development for Managers, Blackwell Publishing

[35] Talbot, C. (1992) Evaluation and Validation: A Mixed Approach, Journal of European Industrial Training, vol. 16, no. 5, pp. 26-32.

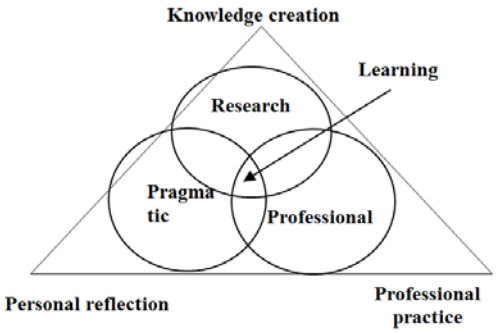
[36] Twitchell, S. (2000) Technical Evaluation Processes in the United States, performance improvement quarterly, p84-110

[37] Tennant (2002) Design of Training Program and Measurement Model, Journal of European Industrial Training, p230-240

[38] Warr, P., Bird, M and Rackham, N (1970) The Evaluation of Management Training, Aldershot, Gower

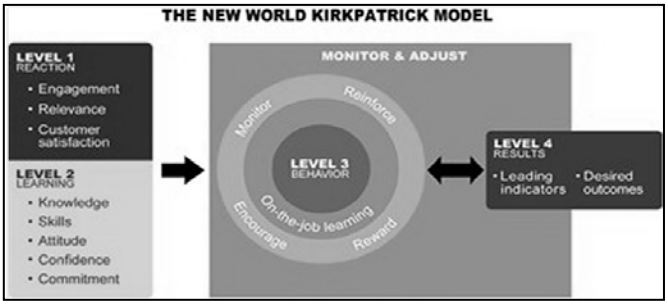
[39] Wang, G. (2003) Valuing learning: the measurement of journey, educational technology, p32-37.

APPENDIX 1: Four purposes of evaluation



Source: Adopted from Sadler-Smith (2006), p385

APPENDIX 2: Kirkpatrick Four levels of evaluation



Source: Kirkpatrick partners (2010-2011)