A CRITICAL ANALYSIS OF A PROGRAM EVALUATION: A CASE STUDY ON THE EFFECTIVENESS OF A TEACHER TRAINING PROGRAM

Chan, K. C. M.
The Hong Kong Polytechnic University
HONG KONG

ABSTRACT

Conducting a program evaluation is a complicated process and it is always not an easy task. A good evaluation would provide useful and important information to different stakeholders including trainers, participants, management, government, professional etc. The process and procedure of selecting appropriate criteria to measure the effectiveness of a training program are often not well defined, particularly what impact of the training program has made in terms of the process of developing knowledge and skills, attitudes towards particular conception, actual behavior change and even more long term impact. This article proposes to evaluate the program effectiveness of a teacher training program offered at one of the University in Hong Kong by using the Kirkpatrick's four-level model (Kirkpatrick 1959a; 1996a). The four level model consists of four evaluation criteria: *reaction*, *learning*, *behavior* and *results*, and each criterion measures different level of impact. The focus of the study is to illustrate the process of how each criterion of the model can be adapted in this context. An overview of evaluation framework, along with examples of evaluation components at each level, are described in order to gain more understanding on the application of the model. Critical analysis is performed on the appropriateness and usefulness of the model when it is being applied. The limitation of the model is also addressed and suggestion on model modification is made for better implementation of the program evaluation of our teacher's training program offered at this University. Although the emphasis of this article is on teacher training program, the ideology and principles of the evaluation framework will be applicable to different levels and types of educational programs.

Keywords: Program evaluation, Higher education, Kirkpatrick model, Teacher training.

INTRODUCTION

Program evaluation plays a vital role in providing feedback on how well of a particular program is offered regardless of any aspects. It is a systematic process of collecting feedback or information on how effective was the program (Goldstein & Ford, 2002) and it should be carried out in a rigorous manner that could be used for future planning and program improvement. A good evaluation would provide useful and important information to different stakeholders, including trainers, participants, management, government, professional etc. Unfortunately, program evaluation is often a complicated process as the procedure of selecting appropriate criteria to measure the effectiveness of the training program are often not well defined. This article proposes to use a popular framework for training program evaluation in many business organizations, Kirkpatrick's four-level model (Kirkpatrick 1959a; 1996a), to evaluate the program effectiveness of a teacher training program offered at one of the University in Hong Kong. The focus of this study is to illustrate the process of how each criterion of the model can be adapted in this context. An overview of the evaluation framework, along with examples of evaluation components at each level, are described so that more understanding can be gained on the application of this particular model. Critical analysis is performed on the appropriateness and usefulness of the model in its application.

The limitation of model is also addressed, with suggestion on model modification for better implementation of how the teacher training program is to be offered in the future.

The Teacher Training Program of the Study

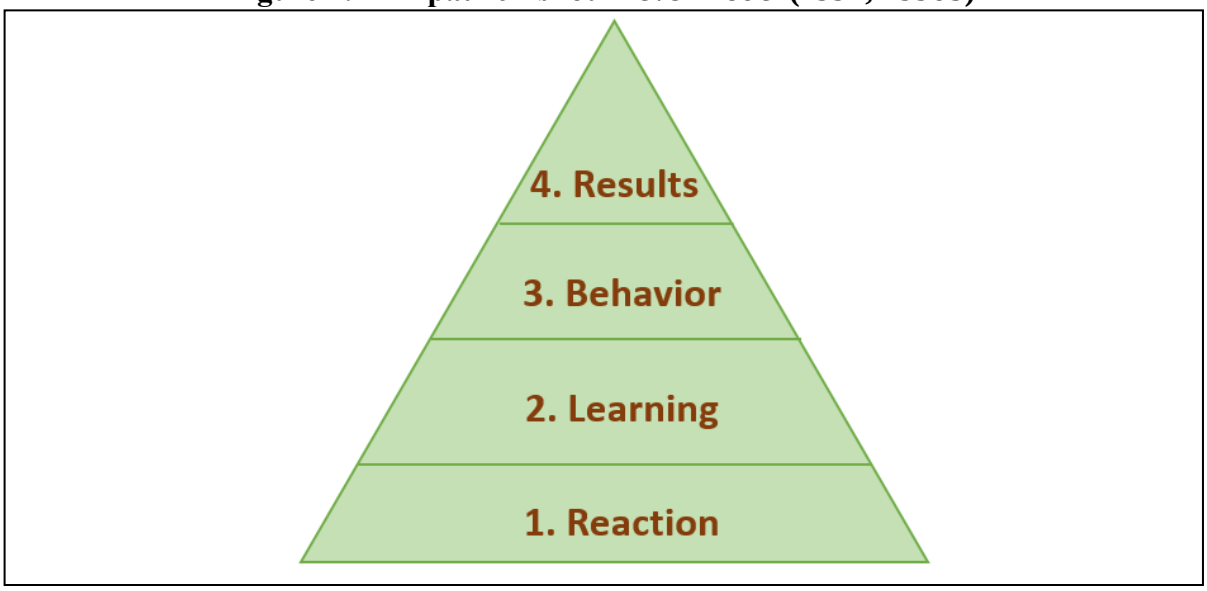
The provision of good quality in teaching and learning is always one of the most important mission of the University. As effective teaching skills play a major role in student learning and their learning outcomes (Timperley, Wilson, Barrar, & Fung, 2008), it is now becoming a common trend to establish initial training to university teachers around the world (Gibbs & Coffey, 2004). As a result, a certificate teacher training program is offered at this University and run by a group of professional in its central unit. The course is a one-week intensive program, which is carried out in a blended mode, consists of face-to-face, online module and teaching practice sessions. It offers twice a year, with class size limited to 30 each intake and it is mandatory to all new teaching staff. At the same time, it is also available for existing teachers who have a felt need to refresh their teaching skills. The program aims to equip teachers with survival skills particularly in their first year of teaching and to enhance teaching effectiveness and student learning. The program covers areas in active learning; student engagement; classroom behavior management; lessons planning; design and use of teaching methods; communication effectiveness; use of technology; choice of assessment methods; improvement on teaching skills and practical session on self-teaching skill demonstration. A reflective module is to be conducted 6 weeks afterward, emphasizing on sharing teaching success and/or obstacles, and identifying challenges faced in participant's teaching. This session serves as an important function for teachers to reflect on how learning from the program can be exercised or transferred into behavior in their classroom teaching.

This teacher training program is one of the most comprehensive program when compared to similar program offered by other local institutions in terms breadth and depth of the topics covered. Each year, approximately 50-60 teachers were participated in this program and two educational development professional are held responsible for planning and conducting this program, with other manpower in coordinating and evaluating the program effectiveness. Given the effort and resources invested into the program, a rigorous program evaluation is used by adapting the Kirkpatrick's four-level model in the hope to provide useful feedback to different stakeholders for future planning and program implementation.

LITERATURE REVIEW

Kirkpatrick's four-level model was developed by Donald Kirkpatrick in 1959, of which it was used for evaluating program effectiveness (Kirkpatrick, 1959a, 1959b, 1960a, 1960b). The framework has stayed relatively unchanged over 40 years even after Kirkpatrick published the book(J. D. Kirkpatrick & Kirkpatrick, 2016). To date, this model is probably the most well-known model and continues to be widely used for assessing training effectiveness (Phillips, 1997a; Arthur Jr, Bennett Jr, Edens, & Bell, 2003; Salas & Cannon-Bowers, 2001). It's wide application has been seen in many sectors including sales and marketing training (Hahne & Schultze, 1996), human performance technology (Rosenberg, 1996), staff training programme in special care unit (Johncox, 2000), medical education (Yardley & Dornan, 2012), programme evaluation in higher education (Praslova, 2010), curriculum intervention assessment (Paull, Whitsed, & Girardi, 2016), Organizational Elements Model for needs assessment and system planning (Kaufman, 1998, cited in Watkins, Leigh, Foshay, & Kaufman, 1998), etc. These four levels are sequential, namely reaction, learning, behavior, and results (see figure 1).

Figure 1: Kirkpatrick's four-level model (1994; 1996b)



Level 1: Reaction. As the word suggests, evaluation at this level measures reaction of participants who took part in the training program. Kirkpatrick refers to "customer satisfaction" or "smile sheet" evaluation, assessing how well training program was perceived. It provides useful information in capturing trainee's feedback about the program (e.g. goals and outcomes) and trainer (e.g. content knowledge, presentation). This level is the easiest level for evaluation and continues to be the most often assessed level (Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997; Arthur et al., 2003; Dysvik & Martinsen, 2008). Reaction evaluation provides fruitful and rich information on program improvement (Antheil & Casper, 1986), which can be used as a standard of performance for future planning (D. L. Kirkpatrick, 1994), 1996b).

Level 2: Learning includes knowledge, skills and/or attitude, which takes place during the training program. For example, what principles and facts were learned; skills and techniques were developed; and/or attitudes were changed during the training program. This level focuses on evaluation of knowledge acquisition and often helps to see if the training program's learning outcomes have been met (Kaufman & Keller, 1994). Shelton & Alliger (1993) suggest that it evaluates the quality of the program and provides support for accomplishing training effects. Evaluation at this level is essential as reaction does not necessary implies learning (e.g. a participant without learning anything could still feel good about the program). Measurement at this level is more difficult and time consuming than previous level, but it helps to identify any behavior change in later stage.

Level 3: Behavior. Evaluation at this level aims to measure actual behavior changes as a result of the training program. In other word, what change in knowledge, skills and/or attitudes can be demonstrated and the extent to which learning is transferred into behavior. Shelton & Alliger (1993) state that learning in previous level can be evident in this level. Kaufman & Keller (1994) state that the focus should be how those newly acquired knowledge and skills operationally applied on the job while Brunken, Plass, & Leutner (2003) describe this level as the ability to 'transfer' the newly acquired knowledge and skills to behavior or new settings. Kirkpatrick stresses the importance of having evaluated levels 1 and 2 before this level as it may help to explain the result of evaluation in this level (e.g. lack of change in

behavior may due to an adverse response in level 1 or a failure in achieving learning objectives in level 2).

Level 4: Results. This level emphasizes on final results or ultimate goals (or organization's goals) of the training program (D. L. Kirkpatrick, 1994). For example, how much did student improved due to training; how much have sales increased as a result of the training. This level often assesses institutional results in terms of its performance improvements, and perceived benefits and costs (Kaufman & Keller, 1994). It provides feedback on the impact and value of training program (Shelton & Alliger, 1993). Kirkpatrick claims that it is the most difficult level for evaluation but probably is the most important level to demonstrate effectiveness or impact of a training program. In some cases, this level can be measured in a relatively easy manner (e.g. sales increase, turnover drop), whereas in many other cases, evaluating this level isn't that straight forward (e.g. training on leadership, communication effectiveness). If results are not that "tangible" and may be affected by many factors, there is a need to seek for positive evidence by relating the results to previous levels (e.g. positive reactions, positive changed in learning and behavior).

Strength and Limitation of The Model

The power of Kirkpatrick's model is the ease of use and it helps evaluators to focus on the understanding of training evaluation in a systematic way (Alliger & Janak, 1989; Shelton & Alliger, 1993). According to Bates (2004), Kirkpatrick's model fits well with the way how training program operates in a real business world as it helps trainers to focus on results that often could align with the organization's objectives. In particular, program evaluation is a complex process and the model helps to simplify such process by focusing on outcomes measurement. Nevertheless, the model has been criticized by many researchers for its assumptions. Firstly, the model assumes that each level is more informative than previous level (Alliger & Janak, 1989). Thus, a measure in learning is perceived to be more informative than a measure in reaction, and so forth. This assumption has been challenged as it does not necessary apply to situation that all four levels can be assessed. Quite often evaluation process only measures criteria in level 1 and level 2 to assess participant's reaction and their learning occur (if any) during the training session (Yusoff et al., 2016).

For example, program for employee rejuvenation may only expected to show an impact at the reaction or learning levels. Secondly, the four levels assume to be causally linked (i.e. the latter is caused by former level), but many past literatures showed weak or no relationship between reactions and other levels (Alliger et al., 1997; Alliger & Janak, 1989). Despite these criticisms, Holton (1996) further argues that Kirkpatrick's model is "a taxonomy of outcome" rather than a model as neither the constructs of the model were fully identified nor empirically validated. Falletta (1998) states that the model is entirely outcome-driven while Brinkerhoff (1987) critiques that the model neglects formative components. The difficulties in evaluating levels 3 and 4 are often mentioned. According to American Society for Training and Development (cited in Watkins et al., 1998), 92% of evaluation studies evaluated at level 1; 34% at level 2; 11% at level 3; and only 2% at level 4. It shows that the complexity of information required may restrain people from an attempt to evaluate the later levels. Although many researchers have further modified and expanded Kirkpatrick's model (Falletta, 1998; Watkins et al., 1998), neither of them are as popular as Kirkpatrick's model. In summary, this model has means of providing practical ideas to professionals rather than purely focusing in scholarly research.

ADAPTATION OF THE FOUR LEVEL MODEL OF EVALUATION CRITERIA TO THE TEACHER TRAINING PROGRAM

This section focuses on the application of Kirkpatrick'sfour-level model on a teacher training program at one of the university in Hong Kong. It illustrates how each criterion of different levels of the model can be adapted in this context for the evaluation of program effectiveness. The rationale of selecting the Kirkpatrick's model based on the questions initiated by the University. These questions were:

1. To what extent participants feel that the teacher training program is useful in helping them to make a start in teaching at the University; how relevant do they perceived the contents of teacher training program to their teaching practices at the University? (level 1)

2. What teaching skills, knowledge and/or attitudes have participants of the teacher training program learned in the program? (level2)

3. How participants' teaching skills, knowledge and/or attitudes improved as a result of the program? (level 3)

4. Have participant's teaching activities had a positive impact on students learning at the University or at wider community? Have students improved in their learning as a result of teachers attending the teacher training program? (level 4)

5. Were the resources spent in the program justifiable? (beyond the model)

These questions (i.e. question #1 to 4)fit well with the evaluation criteria of the Kirkpatrick's model while question #5 addresses another issue beyond the framework of the model. In order to illustrate a better understanding of how to apply the model, Table 1 shows an overall evaluation framework for this teacher training program by adapting the Kirkpatrick's four-level model, with different evaluation components at each level.

Table 1: Evaluation framework for the teacher training program

Evaluation component	Time of data collection	Level 1	Level 2	Level 3	Level 4
Reaction sheet	During; Post	✓			
In-class activities/assessments	During	✓	✓		
Trainer's observation in practical teaching session	Pre-Post		✓	✓	
Self-reported teaching activities	Post			✓	✓
Student feedback questionnaire	Pre-Post; Control group				✓
Approach in Teaching Inventory	Pre-Post				✓

Level 1 - Reaction: This level aims to measure participants' reactions toward the teacher training program in terms of how they feel about the program and what areas of improvement in the program are deemed necessary. This *reaction* criterion includes participant's feeling about the contents of the program; usefulness of materials and examples covered; relevance to teaching practice; perceived increase in teaching skills/techniques; overall satisfaction, etc. and were measured in a reaction sheet. Some reactions of different contents were also captured in the in-class activities. To build on Brinkerhoff (1987)'s perspective about the criticism of Kirkpatrick's model being a summative measure only, assessment in this level

were carried out in both formative and summative process (i.e. measurement takes place during and at the end of program) so that trainers could use this opportunity to make program improvement/modification if needed. Assessment in this level is essentially important to this teacher training program as majority of participants are forced to participate the program at this University. Measuring their responses serve as a vital indicator of perceived usefulness and helpfulness of the program particularly they may be more skeptical in reaction if the participation is made mandatory. Program modification can also be made according to the measure of this level.

Level 2 – Learning: Learning measures the extent to which skills, knowledge and/or attitude takes place during this teacher training program. To evaluate the criteria in this level, learning outcomes of the program were examined. There are numbers of learning outcomes in this program, including techniques to encourage active learning; understanding of our students; preparation for student engagement; methods of engaging students; use of assessment methods; managing classroom behavior; improvement on teaching; sharing good practices and identifying challenges in teaching. Assessments of each of these learning outcomes should be included. According to Kirkpatrick, various types of assessments can be used in this level such as quizzes and tests (multiple choices, true or false), observation and/or checklists, which are all useful in measuring participant's understanding of particular contents delivered by the training program. The challenge is the choice of assessment methods, in which learning can be accurately measured as the result of the training program. In our case study, in-class assessments and observations were used right after learning takes place at each class. Participants were asked to provide examples of how to integrate the techniques learned in the program to their daily teaching practices. Assessments in this level demonstrate how knowledge was received by participants and to be applied in their teaching practices. One most important objective for the teacher training program is the encouragement of actual implementation of good practices in classroom, knowing the extent to which they have learnt in the training program is only a transition of internalizing good teaching practices. More importantly is the degree to which participants can demonstrate their actual performance in next level - behavior.

Level 3 – Behavior: It is always a challenge for teachers to demonstrate how learning of new set of skills/strategies can be transferred to actual behavior in classroom teaching (Showers, 1985) as this level captures behavior changes as a result of the training program. To measure the extent to which behavior changes is demonstrated among participants of the training program, a series of evaluation activities were performed. First, a simple self-reported "post-test" evaluation on changes in teachers' teaching practices were used. This post-test includes questions of actual adoption of the techniques/skills learnt from the training program in classroom teaching (e.g. have they been using active learning skills to engage students in classroom teaching). Participants were also asked to provide actual examples of behavior changes and impact on their teaching practices due to the training program. According to Kirkpatrick, time is required to allow for behavior change to take place. Therefore, the post-test evaluation was carried out 6-8 weeks after participants completed the training program. Second, to move away self-report assessment that may perceive as unreliable and imprecise measure(Spector, 1992), other objective measures such as observation should be used (Verplanken & Orbell, 2003). One challenge of assessment in this level is that it is not easy to encapsulate the change in behavior as an evidence of attending the training program. One way to measure the change of behavior can be demonstrated by a pre-post design. In our example, a teaching practice module is incorporated in the program, in which participants were videotaped for their teaching practices after the training program. To make use of this

opportunity, trainers have also asked participants to prepare one similar video before they attend the training program. Comparison of these "pre-post" videos were against a list of criteria for effective teaching that was introduced in the training program. Thus, any changes in the pre-post design can be attributed to the effect of the training program. Evaluation at this level is extremely useful and important to our program as it is crucial to see if effective teaching components were used by teachers in their teaching practices. This level demonstrates behavior changes in participants' beliefs and/or practices in teaching and therefore it carries very important messages to the university about how teaching is performed at the University. Although Kirkpatrick states that it is not necessary to evaluate all four levels at once, it is certainly, for this teacher training program, to evaluate, at least, up to this level, to demonstrate the effectiveness of the training program.

Level 4 – Results: This level is perceived as the ultimate goal of the training program. In fact it is quite often that assessment at this level relates to the mission of the institution as a whole (Antheil & Casper, 1986) rather than the objectives of any training program per se. In our study, the University aims to provide the training program as a mean to improve teaching effectiveness and to enhance student learning. To measure these objectives, participants were asked to report the impact of their teaching activities on students. To get more reliable measures, two other types of data were used to investigate teaching effectiveness and student learning, they were psychometric Teaching Approach Inventory and Student Feedback Questionnaire. There have been many literatures focusing on how teaching approaches affect students learning. Study showed that there is strong relationships on how teaching impact on students learning and their quality of outcomes (Trigwell & Prosser, 2004). An Approach to Teaching Inventory (ATI) has been widely used in higher education to relate teaching approach (student-focused vs. teacher-focused) to student learning (deep-learning vs. surface-learning). In particular, research showed that the adoption of student-focused approach tended to improve student learning and would lead to better learning outcomes(Ho, Watkins, & Kelly, 2001).

As the training program aims to improve teaching effectiveness and student learning, the Approach to Teaching Inventory were used to investigate if teachers have a higher tendency to adopt a student-focused approach in their teaching as a result of attending the training program. Participants were asked to complete the ATI before and after they attended the training program, any increases in the score on the instrument were used as an evidence of the impact of the training program. Similarly, Student Feedback Questionnaire (SFQ) has been widely used to assess teaching effectiveness although its use is controversial(Olivares, 2003). Research showed that there was an association between student ratings of teaching effectiveness and learning (Arthur Jr, Tubré, Paul, & Edens, 2003) and it was almost certain that SFQ, to certain extent, could provide teachers with views on their teaching practices from student's perspectives and helps to improve teaching effectiveness. Despite the use of pre-post design, Kirkpatrick stressed the significance of using a control group to provide further evidence for impact of any particular training program by comparing the results of "training" group and "no-training" group. In our case, both pre-post design and control group were used to demonstrate such impact. Participant's SFQ results prior attending the training program were used to compare with results after the completion of the training. As some teachers may not have any SFQ results prior the training (e.g. new teachers may attend the program before any teaching is performed), a control group of teachers who haven't attended the training program were used for comparison. Any increase in SFQ ratings were seen as an impact of attending the training program. Measurement at this level is extremely important, particularly in our case study as it demonstrates the impact of training program. Without this

level, there would be much less evidence to demonstrate the ultimate impact of the program. Thus, it is necessary to measure this level even though it requires more resources. The challenge at this level is the difficulty to associate improvement in any of the results (e.g. SFQ and/or ATI score) to the training program as there are many factors influencing student learning and teaching effectiveness. Nevertheless, evaluations at this level can still serve as some evidence of the effectiveness of the training.

Beyond The Model

It is clear that all 4-level of Kirkpatrick's model captured useful information pertaining to measure the impact of teacher training program offered at this University. The model is very helpful for all stakeholders (e.g.management, trainers, teachers) to understanding the impact of how a teaching training program help promoting and improving teaching practices in classroom. There is certainly a need to evaluate all 4-levels for our program, even though many studies assessed only up to level 1 or 2 (e.g. Yusoff et al., 2016). If we don't measure up to level four, evidence of the impact of the training program will not be as strong as we would have demonstrated. However, this model neglects other issue that we have addressed previously, that is, the value and worth of resources put in the training program. Watkins et al. (1998) suggested expanding level 1 of Kirkpatrick's model to include resources availability and process efficiency, which will also help to address our question five (i.e. were the resources spent in the program justifiable?). They have further modified Kirkpatrick framework by adding other elements incorporating societal contributions and tied it to Kaufman's Organizational Elements Model (called "Kirkpatrick Plus"). Other researchers have attempted to extend Kirkpatrick's model including additional level of economic benefit (Hamblin, 1974, cited in Falletta, 1998) and return on investment (Phillips, 1997b). Brinkerhoff (1987) incorporated two formative evaluation levels to compensate the only summative measure of the model. All these extended models suggested that Kirkpatrick's model provides a solid foundation and captures the crucial elements in program evaluation. Nevertheless, the model has its limitation and cannot fully address all of our research questions. As a result, model modification is needed in order to adapt the evaluation need of our training program.

CONCLUSIONS

This study introduces Kirkpatrick's model as a framework for the evaluation tool to investigate the effectiveness of a teacher training program. It demonstrated how Kirkpatrick's model can be implemented on a teacher training program offered at the University. The model helped to bring evidence in identifying impact of our training program at different levels. These four levels provide important and useful information about the effectiveness of the training in teaching practices and particularly helped different stakeholders to understand the concepts and importance of such program. The ease of use of the model encourages all relevant parties to conceptually and practically understand how evaluation activities can be performed to measure program effectiveness. Level 1, with no doubt, is the easiest level for assessment and it is important in our example particularly the participation of the training is mandatory. Although level 2 captures new learning, it is only seen as a process of attaining important outcomes, which happens in later levels. Level 3 and 4 are much more difficult and time consuming to assess, particularly it is not easy to directly associate improvement of teaching effectiveness and its impact on student learning as a result of the training program. The choice of self-reported surveys and other more objective measurements including observation and psychometric inventory helped to increase the credibility of claims and

evidence about the effectiveness of the program. A final remark about the limitation of the model is made regarding to the neglect of resources justification, suggesting that there is a need to extend Kirkpatrick's model framework for our analysis. Although the emphasis of this study is on teacher training program, the ideology and principles of the program evaluation framework will be applicable to different levels and types of educational programs.

REFERENCES

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's Levels of Training Criteria: Thirty Years Later. Personnel Psychology, 42(2), 331–342. http://doi.org/10.1111/j.1744-6570.1989.tb00661.x

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A Meta-Analysis of the Relations Among Training Criteria. Personnel Psychology, 50(2), 341–358. http://doi.org/10.1111/j.1744-6570.1997.tb00911.x

Antheil, J. H., & Casper, I. G. (1986). Comprehensive evaluation model: A tool for the evaluation of nontraditional educational programs. Innovative Higher Education, 11(1), 55–64.

Arthur Jr, W., Bennett Jr, W., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: a meta-analysis of design and evaluation features. Journal of Applied Psychology, 88(2), 234.

Arthur Jr, W., Tubré, T., Paul, D. S., & Edens, P. S. (2003). Teaching effectiveness: The relationship between reaction and learning evaluation criteria. Educational Psychology, 23(3), 275–285.

Bates, R. (2004). A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. Evaluation and Program Planning, 27(3), 341–347. http://doi.org/10.1016/j.evalprogplan.2004.04.011

Brinkerhoff, R. O. (1987). Achieving results from training: How to evaluate human resource development to strengthen programs and increase impact. Jossey-Bass.

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct Measurement of Cognitive Load in Multimedia Learning. Educational Psychologist, 38(1), 53–61. http://doi.org/10.1207/S15326985EP3801_7

Dysvik, A., & Martinsen, Ø. L. (2008). The relationship between trainees' evaluation of teaching and trainee performance among Norwegian executive students. Educational Psychology, 28(7), 747–756.

Falletta, S. V. (1998). Evaluating Training Programs: The Four Levels: Donald L. Kirkpatrick, Berrett-Koehler Publishers, San Francisco, CA, 1996, 229 pp. No longer published by Elsevier.

Gibbs, G., & Coffey, M. (2004). The Impact Of Training Of University Teachers on their Teaching Skills, their Approach to Teaching and the Approach to Learning of their Students. Active Learning in Higher Education, 5(1), 87–100. http://doi.org/10.1177/1469787404040463.

Goldstein, I. L., & Ford, J. K. (2002). Training in organizations: Needs assessment, development, and evaluation Wadsworth. Belmont, CA.

Hahne, C. E., & Schultze, D. E. (1996). Sales and marketing training. The ASTD Training and Development Handbook, 4, 864–884.

Ho, A., Watkins, D., & Kelly, M. (2001). The conceptual change approach to improving teaching and learning: An evaluation of a Hong Kong staff development programme. Higher Education, 42(2), 143–169.

Holton, E. F. (1996). The flawed four-level evaluation model. Human Resource Development Quarterly, 7(1), 5–21. http://doi.org/10.1002/hrdq.3920070103.

Johncox, V. (2000). Evaluability assessment of staff training in special care units for persons with dementia: Strategic issues. Canadian Journal of Program Evaluation, 15(Special issue), 53–66.

Kaufman, R., & Keller, J. M. (1994). Levels of evaluation: beyond Kirkpatrick. Human Resource Development Quarterly, 5(4), 371–380.

Kirkpatrick, D. L. (1994). Evaluating training programs : the four levels (1st ed.). Berrett-Koehler.

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). Kirkpatrick's Four Levels of Training Evaluation. Association for Talent Development.

Olivares, O. J. (2003). A conceptual and analytic critique of student ratings of teachers in the USA with implications for teacher effectiveness and student learning. Teaching in Higher Education, 8(2), 233–245.

Paull, M., Whitsed, C., & Girardi, A. (2016). Applying the Kirkpatrick model: Evaluating an Interaction for Learning Framework curriculum intervention. Issues in Educational Research, 26(3), 490–507.

Phillips, J. J. (1997a). Handbook of training evaluation and measurement methods. Routledge.

Phillips, J. J. (1997b). Measuring return on investment (Vol. 2). American Society for Training and Development.

Praslova, L. (2010). Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in Higher Education. Educational Assessment, Evaluation and Accountability, 22(3), 215–225.

Rosenberg, M. J. (1996). Human performance technology. The ASTD Training and Development Handbook, 384.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. Annual Review of Psychology, 52(1), 471–499.

Shelton, S., & Alliger, G. (1993). Who's afraid of level 4 evaluation? A practical approach. Training and Development, 47(6), 43–46.

Showers, B. (1985). Teachers coaching teachers. Educational Leadership, 42(7), 43–48.

Spector, P. E. (1992). Summated rating scale construction: An introduction. Sage.

The Hong Kong Polytechnic University. (2001). Creating a competitive edge for our students and the community, Strategic Plan for 2001/2 to 2006/7.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2008). Teacher professional learning and development. Retrieved from http://www.orientation94.org/uploaded/MakalatPdf /Manchurat/EdPractices_18.pdf

Trigwell, K., & Prosser, M. (2004). Development and use of the approaches to teaching inventory. Educational Psychology Review, 16(4), 409–424.

Verplanken, B., & Orbell, S. (2003). Reflections on Past Behavior: A Self-Report Index of Habit Strength1. Journal of Applied Social Psychology, 33(6), 1313–1330. http://doi.org/10.1111/j.1559-1816.2003.tb01951.x

Watkins, R., Leigh, D., Foshay, R., & Kaufman, R. (1998). Kirkpatrick plus: Evaluation and continuous improvement with a community focus. Educational Technology Research and Development, 46(4), 90–96.

Yardley, S., & Dornan, T. (2012). Kirkpatrick's levels and education 'evidence'. Medical Education, 46(1), 97–106. http://doi.org/10.1111/j.1365-2923.2011.04076.x.

Yusoff, M. A. M., Ahmad, J., Mansor, A. N., Johari, R., Othman, K., Hassan, N. C., & others. (2016). Evaluation of School Based Assessment Teacher Training Programme. Creative Education, 7(4), 627.