

Kirkpatrick Model: Its Limitations as Used in Higher Education Evaluation

Michael B. Cahapay ^{1,*}

¹College of Education, Mindanao State University, General Santos City, Philippines

ARTICLE HISTORY

Received: May 28, 2020

Revised: Oct. 29, 2020

Accepted: Jan. 02, 2021

Keywords:

Kirkpatrick model,
Program evaluation,
Higher education,
Limitation.

Abstract: One of the widely known evaluation models adapted to education is the Kirkpatrick model. However, this model has limitations when used by evaluators especially in the complex environment of higher education. Addressing the scarcity of a collective effort on discussing these limitations, this review paper aims to present a descriptive analysis of the limitations of the Kirkpatrick evaluation model in the higher education field. Three themes of limitations were found out: propensity towards the use of the lower levels of the model; rigidity which leaves out other essential aspects of the evaluand; and paucity of evidence on the causal chains among the levels. It is suggested that, when employing the Kirkpatrick model in higher education, evaluators should address these limitations by considering more appropriate methods, integrating contextual inputs in the evaluation framework, and establishing causal relationships among the levels. These suggestions to address the limitations of the model are discussed at the end of the study.

1. INTRODUCTION

Evaluation is an essential phase of curriculum and program development in education. Morrison (2003) noted that there are growing pressures to evaluate curriculums and programs in education for different purposes but typically to look into the achievement of the goals. As a result, it can be observed that education borrows evaluation models from other fields like business to evaluate the extent of the achievement of its educational goals. However, the appropriateness of evaluation models is contextually dependent (McNamara, 2000) and the evaluators are faced with the task to adjust them (Fitzpatrick et al., 2004). This is the point where the use of certain evaluation model, not the model itself, presents serious limitations.

Within higher education, one of these models transported to the program evaluation is the model proposed by Donald Kirkpatrick in his seminal articles published in 1959. Historically, the purpose of the Kirkpatrick model was to assist managers for a systematic and efficient means to account for outcomes among employees and in organizational systems. Managers who need solid evidence that training would improve their sales quantity, cost effectiveness, and other business indicators quickly adapted the said model (Yardley & Dornan, 2012).

CONTACT: Michael B. Cahapay ✉ mbcahapay@up.edu.ph 📍 College of Education, Mindanao State University, General Santos City, Philippines.

ISSN-e: 2148-7456 /© IJATE 2021

The Kirkpatrick model originally comprises of four levels - *reaction, learning, behaviour, and impact*. These levels were intentionally designed to appraise the apprenticeship and workplace training (Kirkpatrick, 1976). It is recommended that all programs be evaluated in the progressive levels as resources will allow. Each of these levels have different emphases and are described based on Kirkpatrick & Kirkpatrick (2006):

- The reaction level determines the level of satisfaction of the participants or how they feel about the training program. Assessing how engaged the participant were, how they contributed, and how they responded assists evaluators to recognize how well the participants perceive the training program.
- The learning level measures the level to of knowledge, skills, and values acquired by the participants from the program. This level measures what the participants think they will be able to perform the expected change, how assured they are that they can perform it, and how driven they are to perform it.
- The behaviour level ascertains the changes in the behaviours of the participants in the work environment as a result of the program. The measurement of this level is an activity that should occur over weeks or months following the inputs that the participants received from the training program.
- The impact level examines the institutional outcomes that demonstrate a good return on investment and can be attributed to the training program. Considering the institutional outcomes, a task that can be challenging is to design a method to evaluate these outcomes which are long term in nature.

The general strengths of the Kirkpatrick model in evaluation theory and practice have been extolled by scholars. They recognize the model for its ability to provide the following: simple system or language in dealing with the different outcomes and how information about these outcomes can be obtained; descriptive or evaluative information about the kind of training that are needed, thus allows organizations to anchor the results of what they do in business points of view; and practical approach for the typically complex evaluation process (Bates, 2004). With these strengths, it cannot be denied that Kirkpatrick model has offered significant contributions to the evaluation theory and practice.

Because of the strengths, the Kirkpatrick model has become known in a wide range of evaluation studies. The application of the model has reached the different higher education fields and aspects (see Quintas et al., 2017 on instructional approach; Baskin, 2001 on online group work; Paull et al., 2016 on curriculum intervention; Abdulghani et al., 2014 on research workshops; Aryadoust, 2017 on writing course; Chang & Chen, 2014 on online information literacy course; Farjad, 2012 on training courses; Rouse, 2011 on health information management courses; Dewi & Kartowagiran, 2018 on internship program; Liao & Hsu, 2019 on medical education program; Miller, 2018 on leadership development program; Sahin, 2006 on teacher training program; Masood & Usmani, 2015 on training program; Embi et al., 2017 on blended learning environment).

The reviews of Alliger and Janak (1989), Bates (2004), and Reio et al., (2017) help understand the current state of the Kirkpatrick model by overtly tackling its inherent limitations in the general context. However, an analysis of the limitations when the model is transported to higher education evaluation has not been paid attention. Lambert (2011) supports that judging the worth of learning in the multifarious environments of higher education can be without experiences of limitations. As regards these limitations in the context of higher education, there has been a passing mention (Steele et al., 2016; Covington, 2012; Haupt & Blignaut, 2007) and a collective analysis is yet to be explored.

The intention of this paper is not to downplay the Kirkpatrick evaluation model. It intends to inform evaluators of the possible limitations in the adaptation of such a model in the evaluation in higher education programs or institution. This paper also disclaims that such limitations are not directly attributed to the model. These limitations are based on how the model is applied by evaluators in the educational field. If these limitations are given attention, evaluators will be in a better position as to making cogent considerations to proactively address the potential disadvantages of using the model. As such, they will be guided in designing appropriate methods and tools to successfully use the model and accomplish their desired goals.

Considering the issues and gaps raised in this paper, the current review presents a descriptive analysis of the limitations of the Kirkpatrick model as used in the higher education evaluation.

2. METHOD

This section presents the methods used in this study. It discusses the research design, data sources, data analysis, and analysis procedure. They are elaborated as follows.

2.1. Research design

This research is primarily conducted as a desk review. This research design involves the process of gathering relevant data from various sources (Sileyew, 2019). It may include materials such as legal codes, historical records, statistical data, published papers, news articles, review articles, and other pieces that have a descriptive or analytical purpose (Guptill, 2016). This research design is considered appropriate for this paper. It provides an cogent approach to search, collect, and analyze different materials related to the focus of this paper.

2.2. Data sources

The sources of data for this paper are considered as primary sources. They are original documents, data, or images (Guptill, 2016). These primary sources in the current study consist of books, essays, and articles accessed online. Furthermore, they were screened and included based on the following eligibilities: written in intelligible language, accessible in full text, authored by credible persons or institutions, focused on the Kirkpatrick model as used in higher education evaluation.

2.3. Data analysis

The primary sources gathered in this study were treated through document analysis. It is a technique that “requires repeated review, examination, and interpretation of the data in order to gain empirical knowledge of the construct being studied” (Frey, 2018). Moreover, it involves the creation of themes similar to how interview data are treated (Bowen, 2009). It should be noted, however, that since the themes were readily identified according to the interest of this research, the analysis process was deductively performed (Braun & Clarke, 2006).

2.4. Analysis procedure

The process of deductive analysis was carried out in this study in stages. The researcher initially acquainted himself with the data in the materials, noting down codes relevant to the limitations of the Kirkpatrick model. Then, he grouped these codes based further on the earlier identified themes of limitations of the Kirkpatrick model. The researcher repetitively reviewed the the codes and themes, returning to the original sources until final results were generated.

3. LIMITATIONS OF KIRKPATRICK MODEL AS USED IN HIGHER EDUCATION EVALUATION

This paper is mainly driven by the purpose to provide a descriptive analysis of the limitations of the Kirkpatrick model as it is used by evaluators in the higher education. The following limitations are presented and discussed.

3.1. Propensity towards lower levels of the model

Alliger and Janak (1989) reviewed articles evaluating the Kirkpatrick model. They stated a major conjecture that the levels are structured in increasing order of importance and the model is tiered. Because of this notion, they observed that in the business world, professionals tend to disregard the lower levels of the Kirkpatrick model and address only the higher ones. This is not the case when it comes to higher education.

When the Kirkpatrick model is adapted in educational evaluation, there are pieces of evidence of the tendency to restrict evaluation to the lower levels of the model (Steele, et al., 2016). When it comes to evaluation of effectiveness whether of a training program for teachers or a curriculum for the students in higher education, this limitation can be observed (e.g. see Quintas et al., 2017; Dewi & Kartowagiran, 2018; Sahin, 2006; Aryadoust, 2017). It should be noted that, as disclaimed earlier, these limitations are not caused by the model itself but how it is used in the educational field

Efforts to use the third and fourth levels of the model have been exerted when it comes to evaluation of training effectiveness for teachers in higher education. However, there seemed to be concerns as regards the scope and rigor. For example, in the study conducted by Abdulghani et al. (2014), they evaluated the effectiveness of research workshops to the faculty at a college of medicine. The researchers, however, evaluated the behavioural changes and main outcomes as a single unit in terms of the research activities of the participants. This situation asserts again the limitations as not directly caused by the model itself but how it is used in the field.

Massod and Usmani (2015) also evaluated the outcomes of a training program for teachers in selected medical institutions. The evaluation was framed within the four levels of the Kirkpatrick model. However, the results only discussed the benefits gained by the participants based on their perceptions. These perceived benefits were taken at different points of time to show impacts across the four levels. Moreover, Farjad (2012) attempted a comprehensive evaluation using the Kirkpatrick model in determining the effectiveness of training courses for university employees. The four levels, however, were just measured based on the perceptions of the employees using the survey. The use of perceptions of the participants themselves can be subjective and may decrease the reliability of the results.

The survey of higher education evaluation studies using the Kirkpatrick model in determining the effectiveness of training courses to the employees shows varied evaluation practices. Some studies were restricted on levels one and two. Other studies have tried to reach levels three and four, but they appeared to downplay the scope or diminish the rigor. It should be noted that levels three and four evaluate the workplace behaviours and the organisational impacts respectively (Kirkpatrick & Kirkpatrick, 2006). Rouse (2011) explained that level four operates at the system level or organisational impact. It attempts to identify if an increase in company revenues, client approval, or related indicators is realised as a result of the course or program inputs. Covington (2012) added that while the return of investment is an option to assess economic outcomes, in some professions such as education, optimal outcomes are not exclusively measured by monetary means.

On the other hand, Nickols (2000) explained the propensity towards the lower levels of the Kirkpatrick model in the context of evaluating the impact of the curriculums or programs on the students. He elaborated that any evaluation of change in the student behaviours, level three in the model of Kirkpatrick, will have to occur when they are already in the workplace. It is deemed logical, therefore, to assess behaviour changes in the workplace. However, in the higher education context, employing the Kirkpatrick model can be challenging because students have not normally gone for employment at this stage in their lives.

Hence, because it is difficult to follow the students in the field, many educators tend to end with just the lower levels of the model, leaving out the long-term results of the education. Even if the expectation is clearly defined, it would not be practically easy to trace the learners in the field. Sahin (2006) expressed in a study that an essential limitation when the model is used by evaluators in education is related to the evaluation evidence collected for the behaviour and impact levels. The performance of the students was not directly assessed through observation. Some indirect processes were instead employed to gauge the outcomes of the stated levels.

There are also studies (Embi et al., 2013; Moreira et al., 2018) that attempt to use the level three, but it appears that they also seem to simplify or deviate from the principle of this level. For example, Embi et al. (2013) covered the level three to evaluate transfer of skills in a blended learning environment in higher education. Their result based on student perception showed that students have applied their learning from a direct instruction method into reconstructivist learning. Wang (2018) similarly performed an evaluation study covering the four levels to gauge student learning outcomes as a result of undergoing an information organisation curriculum. Some questionnaire surveys were used specifically to evaluate the behaviour level and results level.

The same can be argued as explained earlier (Kirkpatrick & Kirkpatrick, 2006; Rouse, 2011; Covington, 2012). While the transfer of skills or change in behaviours can be better described through observation, Nickols (2000) reminded that, in using Kirkpatrick model to evaluate the impact of a higher education program to the learners, level three is supposed to measure the medium-term transfer of learned skills from the program to the work environment.

This limitation because of how evaluators use the model in the field may again put barriers and employing the model may be risky for stakeholders especially in education. Thus, the Kirkpatrick model is effectively employed at the lower levels only (Topno, 2012), whether in evaluating training effectiveness to teachers or program effectiveness on students. While higher levels have been attempted to be used in other evaluation efforts, it seemed that the Kirkpatrick model has been treated significantly simplistic. Paull et al. (2016) suggested that similar to challenges experienced in the work, education evaluators should ponder other the alternatives that may be employed to determine the outcomes based on levels three and four.

3.2. Rigidity which leaves out essential aspects of evaluand

The argument for the extreme rigidity of the levels of the Kirkpatrick model is put forward in the light of the importance of contextual factors and essential aspects of the program. This limitation is discussed by various researchers pointing out some features of the Kirkpatrick model with its four-level framework.

For one, according to Rouse (2011), the Kirkpatrick model oversimplifies effectiveness, not considering the various contextual factors within the program. This limitation was also acknowledged in the study of Lillo-Crespo et al. (2017) when they developed a framework adapting the Kirkpatrick model to evaluate the impact of healthcare improvement science. The team noted the weakness of the Kirkpatrick model as devoid of the consideration of contextual influences on the evaluation.

Yardley and Dornan (2012) also observed that, in their study in formal medical education, different levels necessitated different beneficiaries, i.e. levels one to three involve the students; level four relates to the organisations; the educators are overlooked from the system. Thus, they argued that the model does not explore multidimensional outcomes that can be ascertained through qualitative and quantitative approaches. It does not also elaborate on the underlying reasons why outcomes are the outputs of the particular inputs. It appears to gauge only the intended outcomes and disregard the unintended ones.

Furthermore, this problem was echoed by Abernathy (1999). He noted that each level tends to be particular on the questions posed and the outcomes generated, thus rigid. He precluded the levels as not appropriate to evaluate the soft outcomes and continuous education, which are typical in formal education.

3.3. Paucity of evidence on causal chains among the levels

An assumption of the Kirkpatrick model posits that all its levels are contributory (Alliger & Janak, 1989). Grounded on this assumption, scholars and practitioners postulate that, for example, reaction level has a causal influence on learning level. It is believed that the learning level further stimulates change at the behaviour level, and then leads to the desired results at the organisational level (Hilber et al., 1997; Kirkpatrick & Kirkpatrick, 2006).

This assumption can be applied in higher education, that is, what students acquire as a result of participation in the curriculum or program is supposed to cause changes in the reaction, learning, behaviour, and impact. Arthur et al. (2003) determined the relationships among the course grades, student learning, and teaching effectiveness which were reframed within reaction and learning levels. The results revealed that there is a moderate correlation between course grades and student learning. On the other hand, a low correlation was observed between learning measure and teaching effectiveness.

Moreover, Arthur et al. (2010), in their research in the field of technology education, failed to illustrate a piece of evidence of such a relationship between or among the levels. Their findings revealed that substantial relationships between the different levels are restricted. This implies that what long term outcomes learners exhibit might not necessarily be the result of the education they get in school. There could be other external factors that the model does not look into.

This result is empirically supported by Haupt and Blignaut (2007). They applied the Kirkpatrick model in their study to attempted to find out the outcomes in the learning of aesthetics in the program of design and technology education. Similarly, they were not able to show strong corroborations of the causal connections between or among the levels. They specifically were unable to show the link between levels two and three outcomes in their study.

Other related studies previously analysed (Embi et al., 2013; Moreira et al., 2018; Quintas et al., 2017; Dewi & Kartowagiran, 2018; Sahin, 2006; Aryadoust, 2017; Abdulghani et al. 2014; Usmani, 2015; Farjad, 2012; Baskin, 2001; Chang & Chen, 2014) employing Kirkpatrick model did not attempt to probe the causal links among the levels. This concern is not within the interest of these studies.

Tamkin et al. (2002) added that arguably the evaluation model of Kirkpatrick could be negatively attacked on the reasons that empirical studies conducted do not present evidence that the levels are significantly correlated. Hence, it is said to be simple of a thought and that it does not consider other essential features that affect learning. Thus, this limitation should be accounted for when conducting an evaluation using the Kirkpatrick model in higher education, and conclusions should be carefully drawn.

4. CONCLUSION

While Kirkpatrick model is gaining a reputation as a framework for program evaluation, however, it has its limitations in the field of higher education. It presents a propensity towards the use of the lower levels only, rigidity which leaves out other essential aspects of the evaluand, and paucity of evidence on the causal chains among the levels. These limitations offer opportunities and challenges for evaluators who plan to adapt the Kirkpatrick model in higher education evaluation.

First, the propensity towards the lower levels leaves a problem with the limited application of higher levels. This concern may be addressed by considering more appropriate methods and tools. For example, in the behaviour level which seeks to describe how learning has been transferred or has changed the behaviours of the participants in the workplace setting, it is strongly advised that a direct observation must be performed. It should be stressed that an evaluation of change in behaviour requires forceful evidence that goes beyond perceptions of participants usually generated from surveys. Additionally, ultimate outcomes in higher education measured by level four are not exclusively measured by monetary means. Thus, evaluators should redevelop their evaluation frameworks and redesign methods to appropriately evaluate this level.

Furthermore, to offset the argument for too much rigidity of the Kirkpatrick model, a deliberate effort should be made to integrate the contextual inputs and other essential aspects of the evaluand. This can be done by considering the individual participants, work environment, and other aspects that evaluators think are necessary to the framework. For example, a level may be contextualised to educational outcomes or some instruments may be designed to capture these contextual inputs or essential aspects in the light of such limitation. This way, while the levels of the Kirkpatrick model serve as a cogent guide in evaluation, there is room for flexibility so that evaluation will not be too fixed or detached from essential aspects of the evaluand.

Lastly, because the Kirkpatrick model is criticised for the lack of evidence showing the causal relationships among the levels, future studies should strive to prove these chains. The concept of causal relationships can be empirically established through the use of statistical tools. Thus, results in all levels must be converted to a quantitative set as much as possible. Much of the lower levels are often quantitatively measured. Where a level is qualitatively evaluated, data transformation models within the mixed method paradigm offer procedures to convert qualitative data to quantitative data. If causal relationships among the levels are provided attention in evaluation studies using the Kirkpatrick model, more comprehensive and appropriate conclusions may be drawn about the effectiveness of the curriculum or program.

Declaration of Conflicting Interests and Ethics

The author(s) declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author(s).

Authorship contribution statement

Michael B. Cahapay: Conceptualization of the paper, methodology, literature search and evaluation, formal analysis, writing of the manuscript from its first to final form.

ORCID

Michael B. Cahapay  <https://orcid.org/0000-0002-0588-0022>

5. REFERENCES

- Abdulghani, H., A Al Drees, A.M., Khamis, N., & Irshad, M. (2014). Research methodology workshops evaluation using the Kirkpatrick's model: Translating theory into practice. *Medical Teacher*, 36(1), 24-29. <https://doi.org/10.3109/0142159x.2014.886012>
- Abernathy D.J. (1999). Thinking outside the evaluation box. *Training Development*, 53(2), 18-23. <https://eric.ed.gov/?id=EJ578905>
- Alliger, G., & Janak, E.. Kirkpatrick's Levels of Training Criteria: Thirty Years Later. *Personnel Psychology*, 42(2), 331-341. <https://doi.org/10.1111/j.1744-6570.1989.tb00661.x>

- Arthur, W., Jr., Bennet, 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), 234-245. <https://doi.org/10.1037/0021-9010.88.2.234>
- Arthur, W., Jr., Tubre, T. C., Paul, D. S., & Edens, P. S. (2010). Teaching effectiveness: The relationship between reaction and learning criteria. *Educational Psychology*, 23(3), 275-285. <https://doi.org/10.1080/0144341032000060110>
- Aryadoust, V. (2017). Adapting Levels 1 and 2 of Kirkpatrick's model of training evaluation to examine the effectiveness of a tertiary-level writing course. *Pedagogies: An International Journal*, 12(2), 151-179. <https://doi.org/10.1080/1554480X.2016.1242426>
- Bates, R. (2004). A critical analysis of evaluation practice: the Kirkpatrick Model and the principle of beneficence. *Evaluation and Program Planning*, 27, 341-347. <https://doi.org/10.1016/j.evalprogplan.2004.04.011>
- Baskin, C. (2001, December). Using Kirkpatrick's four-level-evaluation model to explore the effectiveness of collaborative online group work. In *Proceedings of the Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education* (pp. 37-44). Melbourne, Australia: Biomedical Multimedia Unit, The University of Melbourne.
- Bowen, G. A. (2009) Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40. <https://doi.org/10.3316/qrj0902027>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Chang, N. & Chen, L. (2014). Evaluating the learning effectiveness of an online information literacy class based on the Kirkpatrick framework. *Libri*, 64(3), 211-223. <https://doi.org/10.1515/libri-2014-0016>
- Covington, J.A. (2012). *Efficacy of webinar training for continuing professional education: applications for school personnel in k-12 settings* [Doctoral dissertation, University of North Carolina] <https://eric.ed.gov/?id=ED550661>
- Dewi. L.R., & Kartowagiran, B. (2018). An evaluation of internship program by using Kirkpatrick evaluation model. *Research and Evaluation in Education*, 4(2), 155-163. <https://doi.org/10.21831/reid.v4i2.22495>
- Embi, Z.C., Neo, T.K., & Neo, M. (2017). Using Kirkpatrick's evaluation model in a multimedia-based blended learning environment. *Journal of Multimedia Information System*, 4(3), 115-122, 2383-7632. <http://dx.doi.org/10.9717/JMIS.2017.4.3.115>
- Farjad, S. (2012). The Evaluation effectiveness of training courses in university by Kirkpatrick model. *Procedia – Social and Behavioral Sciences*, 46, 2837-2841. <https://doi.org/10.1016/j.sbspro.2012.05.573>
- Frey, B. (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation*. SAGE Publications, Inc. <https://doi.org/10.4135/9781506326139>
- Guptill, A. (2016). Secondary sources in their natural habitats. *Writing in College*. <https://milnepublishing.geneseo.edu/writing-in-college-from-competence-to-excellence/chapter/secondary-sources-in-their-natural-habitats/>
- Haupt, G., & Blignaut, S. (2007). Uncovering learning outcomes: explicating obscurity in learning of aesthetics in design and technology education. *International Journal of Technology and Education*, 18(4), 361-374. <https://doi.org/10.1007/s10798-007-9029-1>
- Hilbert, J., Preskill, H., & Russ-Eft, D. (1997). Evaluating training's effectiveness. In L. Bassi & D. Russ-Eft (Eds.), *What works: Assessment, development, and measurement* (pp. 109–150). American Society for Training and Development.
- Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of the American Society of Training and Development*, 13, 3-9.

- Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development* (2nd ed., pp. 301–319). New York: McGraw-Hill.
- Kirkpatrick D. L., & Kirkpatrick J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). Berrett-Koehler Publication.
- Lambert, N. (2011). Ban happy sheets! - Understanding and using evaluation. *Nurse Education Today*, 32(1), 1-4. <https://doi.org/10.1016/j.nedt.2011.05.020>
- Liao, S. C., & Hsu, S. Y. (2019). Evaluating a continuing medical education program: New World Kirkpatrick Model Approach. *International Journal of Management, Economics and Social Sciences*, 8(4), 266-279. <http://dx.doi.org/10.32327/IJMESS/8.4.2019.17>
- Lillo-Crespo, M., Sierras-Davo, M. C., McRae, R., & Rooney. K. (2017). Developing a framework for evaluating the impact of Healthcare Improvement Science Education across Europe: A qualitative study. *Journal of Educational Evaluation in Health Profession*, 14, 28. <https://doi.org/10.3352/jeehp.2017.14.28>
- Masood, R., & Usmani, M. A. W. (2015). A study for program evaluation through Kirkpatrick's model. *Khyber Medical University Journal*, 2(7), 76-80. <https://www.kmu.edu.pk/article/view/15377>
- Miller, B.J. (2018). *Utilizing the Kirkpatrick model to evaluate a collegiate high-impact leadership development program* [Master's thesis, Texas A&M University, College Station, Texas]. <https://oaktrust.library.tamu.edu>
- Moreira I.C., Ramos, I., Ventura, S.R., Rodrigues, P.P. (2018). Learner's perception, knowledge and behaviour assessment within a breast imaging E-Learning course for radiographers. *European Journal of Radiology*, 111, 47-55. <https://doi.org/10.1016/j.ejrad.2018.12.006>
- Morrison J. (2003). ABC of learning and teaching in medicine: Evaluation. *British Medical Journal*, 326(7385), 385-387. <https://doi.org/10.1136/bmj.326.7385.385>
- Nickols, F. W. (2000). *Evaluating training: There is no "cookbook" approach*. http://home.att.net/~nickols/evaluating_training.htm
- 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), 490-502. <https://www.iier.org.au/iier26/paull.pdf>
- Quintas, C., Fernandes Silva, I., & Tiexiera, A. (2017). Assessing an e-Learning and b-Learning Model - A study of perceived satisfaction. *International Journal of Information and Education Technology*, 7(4), 265-268. <https://doi.org/10.18178/ijiet.2017.7.4.878>
- Reio, T. G., Rocco, T. S., Smith, D. H., & Chang, E. (2017). A Critique of Kirkpatrick's Evaluation Model. *New Horizons in Adult Education & Human Resource Development* 29(2), 35-53. <https://doi.org/10.1002/nha3.20178>
- Rouse D. N. (2011). Employing Kirkpatrick's evaluation framework to determine the effectiveness of health information management courses and programs. *Perspectives of Health Information Management*, 8, 1-5. <https://www.ncbi.nlm.nih.gov/pubmed/21464860>
- Sahin, V. (2006). Evaluation of the in-service teacher training program "The Certificate for Teachers of English" at the Middle East Technical University School of Foreign Languages. [Doctoral dissertation, Middle East Technical University] <https://etd.lib.metu.edu.tr/upload/12607752/index.pdf>
- Shelton, S., & Alliger, G. M. (1993). Who's afraid of level 4 evaluation? A practical approach. *Training and Development Journal*, 47, 43-46. <https://eric.ed.gov/?id=EJ463549>
- Sileyew, K. J. (2019). Research design and methodology. In E. Abu- Taieh, A. El Mouatasim, & I.H. Al Hadid (Eds.), *Cyberspace*. IntechOpen. <https://doi.org/10.5772/intechopen.78887>

- Steele, L. M., Mulhearn, T. J., Medeiros, K. E., Watts, L. L., Connelly, S., & Mumford, M. D. (2016). How do we know what works? A review and critique of current practices in ethics training evaluation. *Accountability in Research*, 23(6), 319-350. <http://dx.doi.org/10.1080/08989621.2016.1186547>
- Tamkin, P., Yarnall, J., & Kerrin, M. (2002). *Kirkpatrick and beyond: A review of models of training evaluation* (Report No. 392). Institute for Employment Studies
- Topno, H. (2012). Evaluation of training and development: An analysis of various models. *IOSR Journal of Business and Management*, 5(2), 16-22. <https://doi.org/10.9790/487x-0521622>
- Yardley, S., & Dornan, T. (2012). Kirkpatrick's levels and education 'evidence'. *Medical Education*, 46(1), 97-106. <https://doi.org/10.1111/j.1365-2923.2011.04076.x>