

Kirkpatrick Model and Training Effectiveness: A Meta-Analysis 1982 To 2021

Fahad Nawaz¹, Wisal Ahmed², Muhammad Khushnood³

Abstract

By examining the overall success of managerial training concerning Kirkpatrick's training effectiveness paradigm, this work seeks to contribute to the substantial contributions of prior 40-year research in the area. Additionally, this study seeks to assess the overall findings regarding its renowned levels, reaction, learning, behavior, and results of Kirkpatrick's training model and associations among these levels. Through a meta-analytic process, this study statistically extends and unifies the management training literature. The Kirkpatrick model was the subject of a meta-analysis that covered 41 papers (n=41) between 1982 and 2021. Although accommodating literary study regarding Kirkpatrick's four levels of the training assessing model recommended positive association among its distinct levels, the results do not indicate a significant development in the usefulness of managerial training from 1982 through 2021. The implications have a direct bearing on the choice of evaluation techniques for upcoming research on the effectiveness of management training programs. The academic world and practitioners both value this implication. The potential exclusion of prior research and the variety of assessment techniques employed in earlier studies—beyond the simple categories of objective and subjective assessment—are among the study's limitations. The fact that this study spans a significant amount of time is its key contribution. The approach thus provides a wider perspective on managerial training throughout time.

Keywords: Kirkpatrick Model, Meta-Analysis, Reaction, Learning, Behavior, Results

1. Introduction

Human resource development efforts concentrate on skill and knowledge enhancement of their workforce. The training and individual capacity building activities improve the workers' innate abilities, knowledge, and performance outcomes (Dachner,

1 PhD Scholar, Institute of Business Studies, Kohat University of Science & Technology.

Email: fahad.nawaz1@gmail.com

2 Professor, Institute of Business Studies, Kohat University of Science & Technology.

Email: dr.wisal@kust.edu.pk

3 Associate Professor, Institute of Business Studies, Kohat University of Science & Technology.

Email: mkhushnood@kust.edu.pk

ARTICLE HISTORY

07 May, 2022 Submission Received

16 Jun, 2022 First Review

02 Aug, 2022 Second Review

05 Oct, 2022 Accepted

Ellingson, Noe & Saxton, 2021). Therefore, without having well trained employees the organization will not be able to accomplish organizational goals (Samwel, 2018). Consequently, the companies spend thousands of dollars on training activities (Patki, Sankhe, Jawwad & Mulla, 2021). According to the United States Industrial training report (2020), global firms invest \$696.7 billion on training activities. Similarly, Asian countries are also spending a significant sum on training and education technical education. The existing training evaluating models are assessing the training effectiveness without an appropriate mechanism (Hazan-Liran, & Miller, 2020; Velada & Caetano, 2007). Kirkpatrick introduced a training evaluation model in 1960. According to Cahapay (2021), the Kirkpatrick's paradigm was designed through an effective and productive technique to assess learning outcomes among individuals and organizational structures concerning training.

There are four levels in the Kirkpatrick Training Evaluation Model (KTEM); namely, trainee reaction, learning, behavior, and result (Ho, Arendt, Zheng & Hanisch, 2016). The Kirkpatrick model revealed substantial correlations between the four stages of training effectiveness. However, only a small number of research studies have strongly validated these linkages (Alsalamah & Callinan, 2021; Manzoor & Din, 2019). Research scholars reported that the organizations frequently neglect to assess the behavior and result of the training effectiveness paradigm due to the challenges involved in its assessment (Alliger & Janak, 1989; Clement, 1982; Homklin, 2014). Moreover, it was also reported that during assessing the behavior, and result, the participants' responses were quite biased (Abdelhakim et al., 2018).

In spite of the fact that Kirkpatrick model has established the significant interconnections among the four levels of training effectiveness, but very limited studies have substantiated these relationships empirically (Alsalamah & Callinan, 2021; Baluku, Matagi & Otto, 2020; Costabella, 2017). The Kirkpatrick approach is also mentioned in numerous literatures, including research articles, novels, conference papers, and gray literature, as per the literature review conducted for this study. Additionally, it was shown that there aren't enough studies worldwide that use the Kirkpatrick model in this domain of social sciences.

The meta-analysis review has found that there are currently 298 research articles, 20 case studies, 48 conference papers, nine books, 49 reviews, one short survey, and 48 books associated with the Kirkpatrick paradigm in general (all fields). On the other hand, there is only a single case study, sixteen conference papers, three books, twenty-six reviews, no brief survey, and 123 research articles published up to this point in the social sciences. The illustration shows how few papers linked to the Kirkpatrick model were between 2011 and 2021. To sum up the discussion, up till now three gaps are identified. First, mix findings discovered pertaining to the association between

four levels of Kirkpatrick model (Alsalamah & Callinan, 2021; Baluku, Matagi & Otto, 2020; Costabella, 2017). Second, ignorance in measuring the level three, i.e., behavior and level four, i.e. result (Alliger & Janak, 1989; Clement, 1982; Homklin, 2014) and lastly, the existence of little studies pertaining to the Kirkpatrick model particularly, in the social sciences domain.

Therefore, to bridge the current research gap, it is necessary to do a systematic review (meta-analysis) of the four levels of the Kirkpatrick model. Here the focus on evaluating the association between reaction towards learning, learning towards behavior, and behavior towards the result. Therefore, the study objective is to do systematic review of the relationship among four levels of the Kirkpatrick model. This study offers the accommodative literary work regarding Kirkpatrick's four levels of the training effectiveness. The study is focused to see the causal association between KTEM; in the prior literature. Particularly, the study enhances the understanding of causal association among four levels. The study assessments might be fruitful to enhance the literature of the training modules and will propose the suitable human resource development (HRD) strategies for companies. Moreover, the study may deliver valuable knowledge of training effectiveness and the baseline for training evaluation to academia and industry.

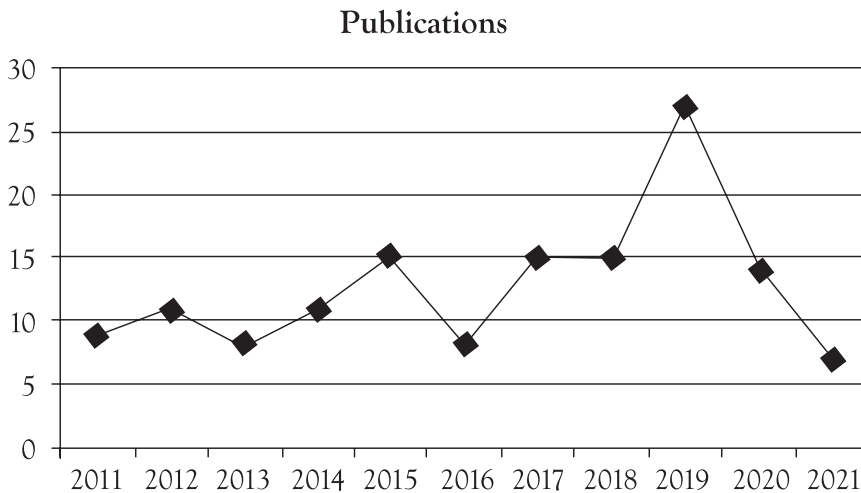


Figure 1: Published Studies (2011-2021)

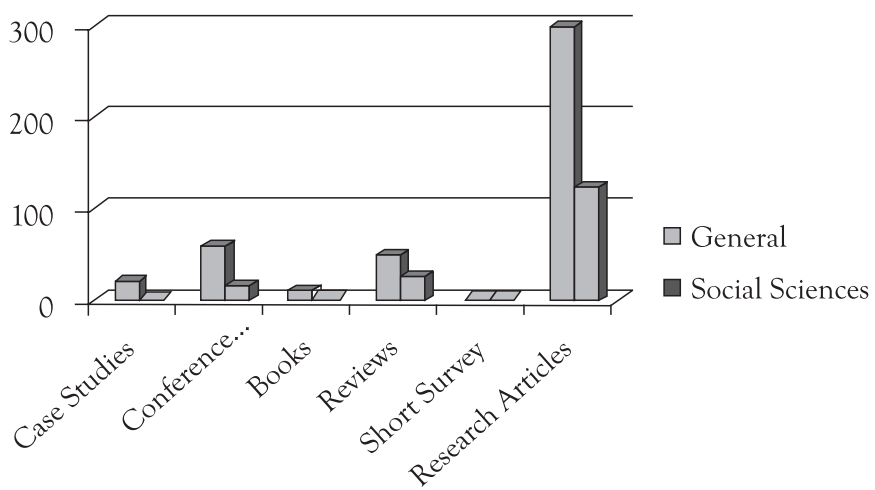


Figure:2 Kirkpatrick Literature in all Fields

2. Kirkpatrick Model-Meta-Analysis

The author performed a meta-analysis of studies on training effectiveness using the Kirkpatrick model to examine the paradigm thoroughly. To conduct a systematic review in the framework of the Kirkpatrick training, assessing model from 1982 to 2021, a total of 41 studies (n=41) was considered through a total sample (n=8825). Thirty-three studies (n=33) had been associated with the trainee reaction (level-one) and trainee learning (level-two), twenty-nine studies (n=29) were associated with trainee learning (level-two) and trainee behavior (level-three), and just three studies (n=3) were related to trainee behavior (level-three) and trainee result (level-four). The following table provides a detailed breakdown of the journals, authors, and countries used as the research and a systematic review foundation.

Table 1: Studies Selected for Meta-Analysis

S#	Authors	Journal	N	Country
1	Clement (1982)	Public Personnel Management	50	USA
2	Wexley & Baldwin (1986)	Academy of Management Journal	120	USA
3	Baldwin (1992)	Journal of Applied Psychology	72	USA
4	Warr & Bunce (1995)	Personnel Psychology	106	USA
5	Cannon-Bowers et al. (1995)	Military Psychology	1037	USA
6	McEvoy (1997)	Society of Human Resources Mgt	140	USA
7	Fisher & Ford (1998)	Personnel Psychology	121	USA

8	Warr, Allan & Birdi (1999)	Journal of Occup'l & Org'l Psychology	163	UK
9	Bates et al. (2000)	Human Resource Development Int'l	150	USA
10	Frayne & Geringer (2000)	Journal of Applied Psychology	30	USA
11	Tracey et al. (2001)	Human Resource Development Quarterly	420	USA
12	Richman-Hisrich (2001)	Human Resource Development Quarterly	1335	USA
13	Gully et al. (2002)	Journal of Applied Psychology	181	USA
14	Tan, Hall & Boyce (2003)	Human Resource Development Quarterly	283	USA
15	Liao & Tai (2006)	Social Behavior Personality	132	USA
16	Savoldelli et al. (2006)		Anesthe- siology	42
17	Lim, Lee & Nam (2007)	Int'l Journal of Information Mgt	170	Japan
18	Sulsky & Kline (2007)	Int'l Journal of Training & Devel- opment	65	Canada
19	Bell & Ford (2007)	Human Resource Development Quarterly	113	USA
20	Liebermann & Hoffman (2008)	Int'l Journal of Training & Devel- opment	213	Germany
21	Sitzmann et al. (2009)	Academy of Management Pro- ceedings	125	USA
22	Orvis et al. (2009)	Journal of Applied Psychology	274	USA
23	Welke et al. (2009)	Anesthesia and Analgesia	30	Canada
24	Grant et al. (2010)	Clinical Simulation in Nursing	40	USA
25	Fisher et al. (2010)	Journal of Applied Psychology	237	USA
26	Van Heukelom et al. (2010)	Simulation in Healthcare	161	USA
27	Lin, Chen & Chuang (2011)	Int'l Journal of Management	494	Japan
28	Boet et al. (2011)	Critical Care Medicine	50	Canada
29	Shinnick et al. (2011)	Clinical Simulation in Nursing	168	USA
30	Saks & Burke (2012)	Int'l Journal of Training & Devel- opment	150	Canada
31	Dreifuerst (2012)	Journal of Nursing Education	238	USA
32	Chronister & Brown (2012)	Clinical Simulation in Nursing	60	USA
33	Reed et al. (2013)	Clinical Simulation in Nursing	64	USA

34	Mariani et al. (2013)	Clinical Simulation in Nursing	86	USA
35	Homklin (2014)	Int'l Journal of Training & Development	228	Thailand
36	Grant et al. (2014)	Nurse Education in Practice	48	USA
37	Reed (2015)	Nurse Education in Practice	58	USA
38	Weaver (2015)	Clinical Simulation in Nursing	96	USA
39	Liao & Hsu (2019)	Int'l Journal of Mgt, Economics & SS	393	Japan
40	Manzoor & Din (2019)	Journal of Managerial Sciences	732	Pakistan
41	Zielilska-Tomeczak et al. (2021)	Nutrients	150	Switzerland

Note. Meta-Analysis Studies

2.1 PRISMA Model

The PRISMA (preferred reporting item systematic review and meta-analysis) was used to carry out the systematic review (Moher et al., 2015). Page et al.(2021) suggested PRISMA model for meta analytical review. The model is used for several reasons a) PRISMA model aims to help authors improve the reporting of systematic reviews, b) The PRISMA flow diagram visually summarizes the screening process, and c) The PRISMA model is relevant for mixed-methods systematic reviews which include quantitative and qualitative studies (Moher et al., 2015). The PRISMA consisted of four parts, i.e., identification, screening, eligibility, and inclusion of studies. The PRISMA is recognized as standard for reporting evidence in systematic reviews and meta-analyses. The PRISMA a) demonstrate quality of review, b) allows readers to assess weakness and strengths, c) allow replications of review and d) structure and format the review (Moher et al., 2015).

2.1.1 Rational of Using PRISMA Model

In this study the researchers have used the PRISMA model for several reasons. First, PRISMA model describe the contemporary state of knowledge, understanding and relevant uncertainties (Sampson, Tetzlaff & Urquhart, 2011). Second, the PRISMA model coherence the significance of the review (Deeks, 2002). Third, PRISMA model assist the scholar to enhance the meta-analytical review (Hoffmann et al., 2017). Fourth, PRISMA may also be useful for critical appraisal of published systematic reviews, although it is not a quality assessment instrument to gauge the quality of a systematic review (Sampson, Tetzlaff & Urquhart, 2011). Lastly, PRISMA allows and reports the effort of intervention about the variables in prior literature (Moher et al., 2015).

3. Selection Criteria (Study)

The studies were selected from the year 1982–2021. In the identification phase a sum of (n=883) studies were found, i.e., (n=774) studies from database search, and (n=109) extra records through other means. Out of (n=883) studies, (n=316) studies were removed due to replica and (n=567) studies were screened out. In screening phase, out of (n=567) studies about (n=443) articles had been left out due to record replications and (n=124) studies were screened out and found eligible. During the eligibility phase, (n=33) studies were eliminated because abstract not matched with the study variables. In the inclusion phase out of (n=91) studies (n=50) studies were not included because the studies were based on qualitative viewpoints. Finally, (n=41) quantitative studies were included to conduct the meta-analysis. The PRISMA model figurative representation is then described.

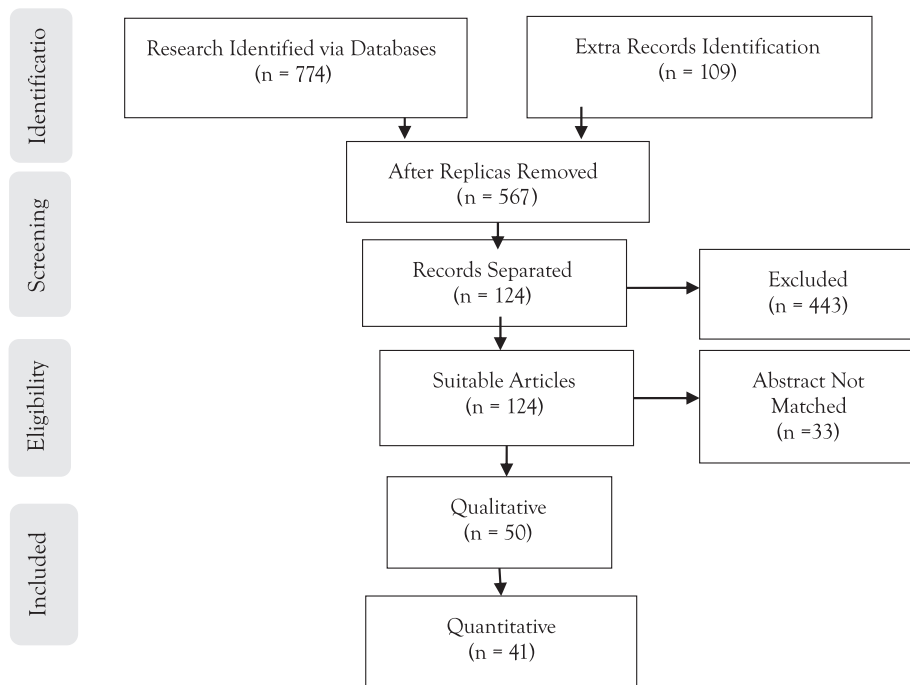


Figure 3: PRISMA Flow Diagram for Meta-Analysis

4. Methodology

The research philosophy was positivism and the systematic review was analyzed via meta-analysis based on the studies related to the quantitative nature. The PRISMA (preferred reporting item systematic review and meta-analysis) model was used to carry out the systematic review (Moher et al., 2015). The Web of Science (WoS), Scopus,

and Science Direct are the search engines (Balstad & Berg, 2020) which were selected as sources for searching the papers. The exclusion of Google Scholar was due to several factors. First, compared to many other archives like WoS or Scopus, its indexing practices are less stringent, often resulting in a less effective search results (Shareefa & Moosa, 2020). Furthermore, the findings cannot be extracted, in contrast to the majority of other resources like Science Direct and WoS (Moral-Munoz et al., 2020). So, WoS, Scopus, and Science Direct were the sole three search resources used by the investigator. The research adopted three strategies to locate pertinent research publications. To start, the investigator accessed three internet database systems: WoS, Scopus, and Science Direct. There are no date constraints because the date was set to default. Title, Abstract, and Phrases were prioritized in the search parameters for many indexing sources.

The investigator looked through the records up until 31st December 2021. Second, the review searches internet resources for ancestors-related existing publications and studies to find forebears. Third, the researcher assessed the descendancy of journals that cited publications using the Kirkpatrick Model. N was the number of the sample group for this inquiry. The development of this encoded framework is a component of a more thorough meta-analysis and its relationships. By using Fisher z (hyperbolic arctangent) transformation ($z \tanh^{-1}(r)$) to analyze a Pearson correlation, the researcher then used Steiger (1980) methods to compute the variance and covariance of z -transformed values by using Hedges and Olkin (1985) meta-analysis procedure. The investigator acquired and transformed the data sequentially into Pearson correlation (r) (Schulze, 2004).

Researchers used random-effects test to analyze data as either a subset of a heterogeneous population from which they meant to draw inferences or even as a whole group from which they hoped to draw generalizations (Borenstein et al., 2010). The researcher fitted a random-effects model using the maximum-likelihood method and the JASP software. The author analyzed the variety of direct population results and provided reliability and accuracy ranges since they anticipated heterogeneity in effect magnitude. The accuracy of the parameters calculated is reflected in the field of possibilities within which researchers can be confident that the underlying mean of the responses lies. The range in which the majority of path coefficients lie, or the confidence intervals, show the variety of influence sizes for a population (Whitener, 1990). To achieve the highest level of precision in search queries, the terms used across the question were selected and determined by several characteristics. This study focuses on the Kirkpatrick model and how it could be used as an evaluation system. These terms were part of the investigation since the Kirkpatrick model is “a paradigm, structure, framework, typology, approach, and typographic (Holton, 1996,

p. 50). The investigator used a database, keyword search, tools, coding, or analysis procedures to do a meta-analysis, as stated in Table 2.

Table 2: Method (Meta-Analysis)

S#	Description	Instrument Used
1	Analysis	Correlation Pearson
		Random Effects Method
		Reaction (A)
		Learning (B)
2	Coding	Behavior (C)
3	Software	Result (D)
		JASP
4	Databases Used	Scopus
		Science Direct
		Web of Knowledge

Note. Meta-Analysis Methods

4.1 Reaction to Learning

Thirty-three investigations (n=33) in total were calculated using a forest plot to determine the association between trainee reaction (level-1) and learning (level-2). The forest plot is a visual depiction of results from various research investigations focusing on the same topic and their combined estimate (Lalkhen & McCluskey, 2008). Two columns display the forest plot. The studies' names are listed in the left column, usually in chronological sequence from top to bottom. Confidence intervals are shown as horizontal lines in the figure in the right column that displays the odds ratio measurement for each of these investigations. A vertical line that denotes no effects also was evident. This line will be parallel to a range for independent research if there is no influence at the point estimate. The same would be true for such an influence gauge that was the subject of the meta-analysis. If the diamond's vertices cross the lines of no effects, the conclusion of the meta-analysis cannot be deemed to differ from no product at the specified confidence level. Out of 33 investigations (n=33), 30 (n=30) studies confirmed an excellent relationship between trainee reaction (level-1) and learning (level-2), according to the forest plot data. However, only three (n=3) investigations found a negative correlation between level-one trainee reaction and learning (level-2). Due to the random effect estimation, the total value represents the actual observed outcome of all studies, which is on the right and somewhat more than zero ($r=.23$, CI [.07,.39]). The diamond at the bottom shows this. This suggests

that trainee reaction (level one) and learning have a good relationship (level two). Additionally, the overall observed result across all investigations is on the right-hand side and slightly above zero, indicating a positive correlation between trainee reaction and learning (level-1) (level-2).

4.2 Learning to Behavior

The link between trainee learning (level 2) and behavior (level 3) was assessed using a forest plot over a total of 29 experiments ($n=29$). Out of twenty-nine research ($n=29$), according to forest plot data, it is discovered that twenty-six ($n=26$) studies verified a good correlation between trainee learning (level-2) and behavior (level-3). Only three researches ($n=3$) revealed a negative correlation between trainee learning (level 2) and behavior (level 3). Based on the random effect estimation, the total value shows the actual observed outcome of all studies, which is on the right and somewhat more than zero ($r=.28$, CI $[.17,.38]$). The diamond at the bottom shows this. This suggests that trainee learning (level-2) and behavior (level-3) positively correlated. Additionally, the overall observed result across all studies is on the right-hand side and slightly above zero, indicating a positive correlation between trainee learning and behavior (level-2) (level-3).

4.3 Behavior to Result

In three studies, a total of $n = 3$ was calculated using a forest plot to determine the association between trainee behavior (level-3) and result (level-4). According to forest plot results, all three ($n=3$) investigations indicated a strong correlation between trainee behavior (level-3) and result (level-4). Based on the random effect estimation, the total value shows the actual observed outcome of all studies, which is on the right and somewhat more than zero ($r=.44$, CI $[.27,.1.15]$). The diamond at the bottom shows this. This indicates that there is generally a good correlation between trainee behavior (level 3) and result (level 4). Additionally, the overall observed outcome of the result is slightly above zero on the right-hand side, indicating a positive connection between trainee behavior (level-3) and result (level-4).

4.4 Summary of Meta-Analysis

A comprehensive sample ($n=8825$) of papers ($n=41$) about the Kirkpatrick model from 1982 to 2021 was considered in the meta-analytical review. Thirty-three ($n=33$) of the forty-one studies had been concerned with trainee reaction (level-1) and learning (level-2), whereas twenty-nine ($n=29$) were concerned with trainee learning (level-2) and behavior (level-3), and just three had been involved with trainee behavior (level-3) and result (level-4). The standard procedure, known as PRISMA, was used to conduct

the systematic review. The identification, selection, eligibility, and inclusion of studies were all parts of this procedure. As of 1982–2021, approximately $n=774$ studies were found overall through evaluation and database searches, and an extra ($n=109$) record were found from other sources (the Kirkpatrick model). The Web of Science, Scopus, and Science Direct were selected as repositories for articles pertinent to the study's topic. The researcher fitted the random-effects model using the maximum likelihood estimation and the JASP software. According to forest plot data, thirty ($n=30$) research out of thirty-three papers ($n=33$) confirmed a connection between trainee reaction (level-1) and learning (level-2). Comparatively, only three research ($n=3$) found a negative correlation between trainee reaction (level-1) and learning (level-2). Second, 26 studies out of the twenty-nine papers ($n=29$) indicated that there is a positive correlation between trainee learning (level-2) and behavior (level-3). Comparatively, only three research ($n=3$) revealed a negative correlation between trainee learning (level-2) and behavior (level-3). Third, it is discovered that all three ($n=3$) investigations from the three papers ($n=3$) confirmed a positive correlation between trainee behavior (level-3) and trainee result (level-4). The Figure (4, 5 and 6) shows the association between level-1, i.e., a) reaction to learning, level-2, i.e., b) learning to behavior and level-3, i.e., c) behavior to result. The meta-analysis review's summary is shown in Table 2.4.

Table 2.4: Summary (Meta-Analytic Review)

S#	Variables Relationship	Positive Relationship	Negative Relationship
1	Reaction (Level-1) and Learning (Level-2)	30	3
2	Learning (Level-2) and Behavior (Level-3)	26	3
3	Behavior (Level-3) and Result (Level-4)	3	0
4	Total Relationships Identified	56	6

5. Discussion

Numerous training, assessment studies have failed to identify obvious causal correlations among four levels of training evaluating model (Alliger et al., 1997; Alliger & Janak, 1989). The sequential ordering of training effectiveness has, though, rarely been studied in training and assessment (Alliger et al., 1997). On the other hand, only a small number of training investigations have shown some evidence in favor of the hierarchical order correlation of all four stages (Liao & Hsu, 2019; Manzoor & Din, 2019; Homklin, 2014; Saks & Burke, 2012; Alliger & Janak, 1989). According to literature, there exist a mixed-findings pertaining to the Kirkpatrick four level re-

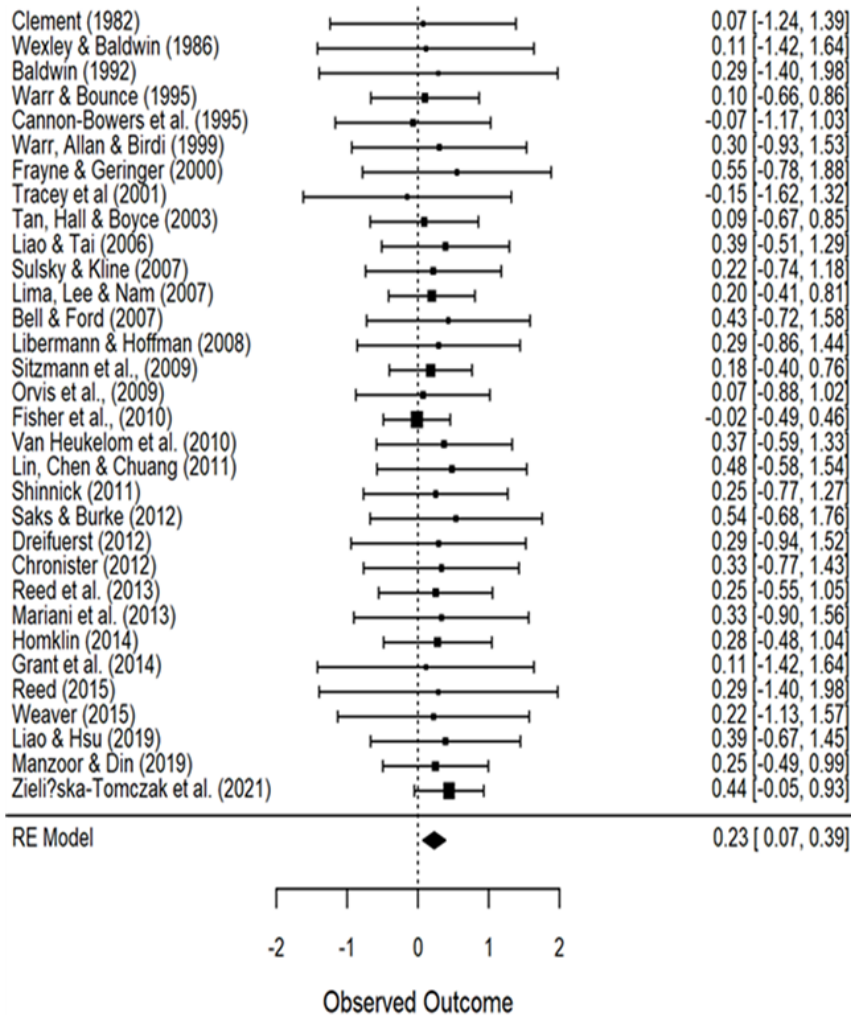


Figure 4: Forest Plot Outcome (Reaction to Learning)

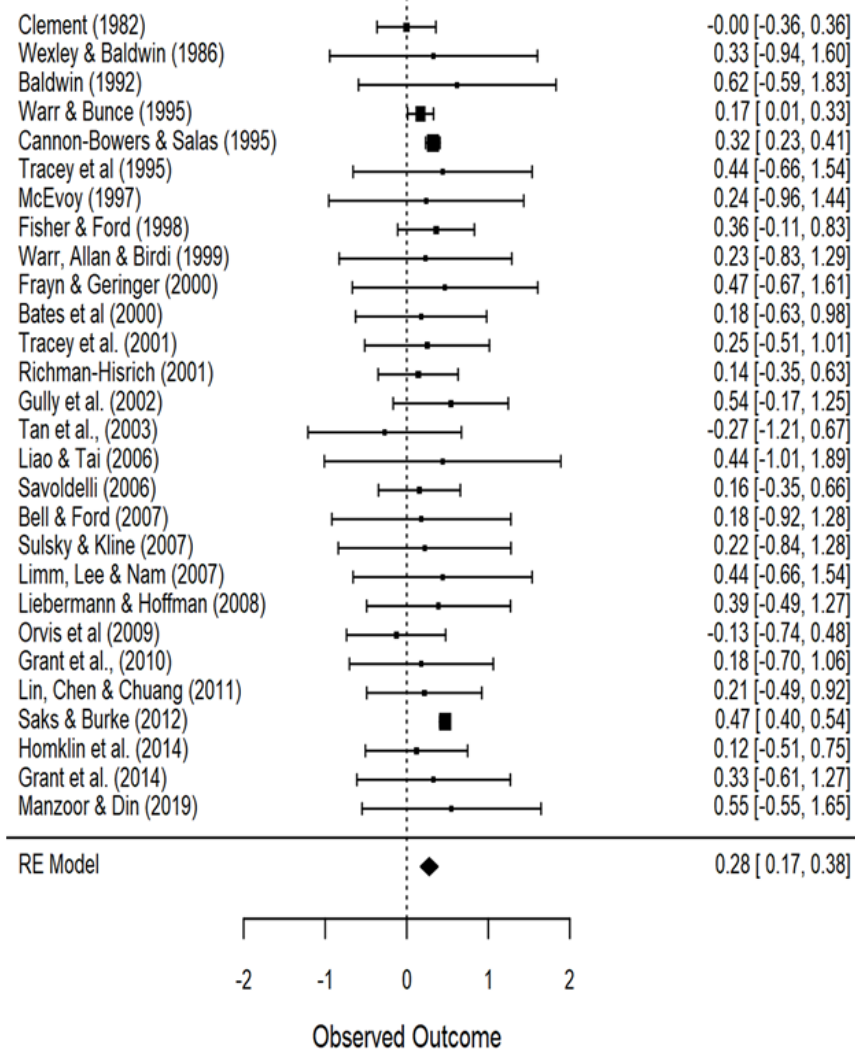


Figure 5: Forest Plot Outcome (Learning to Behavior)

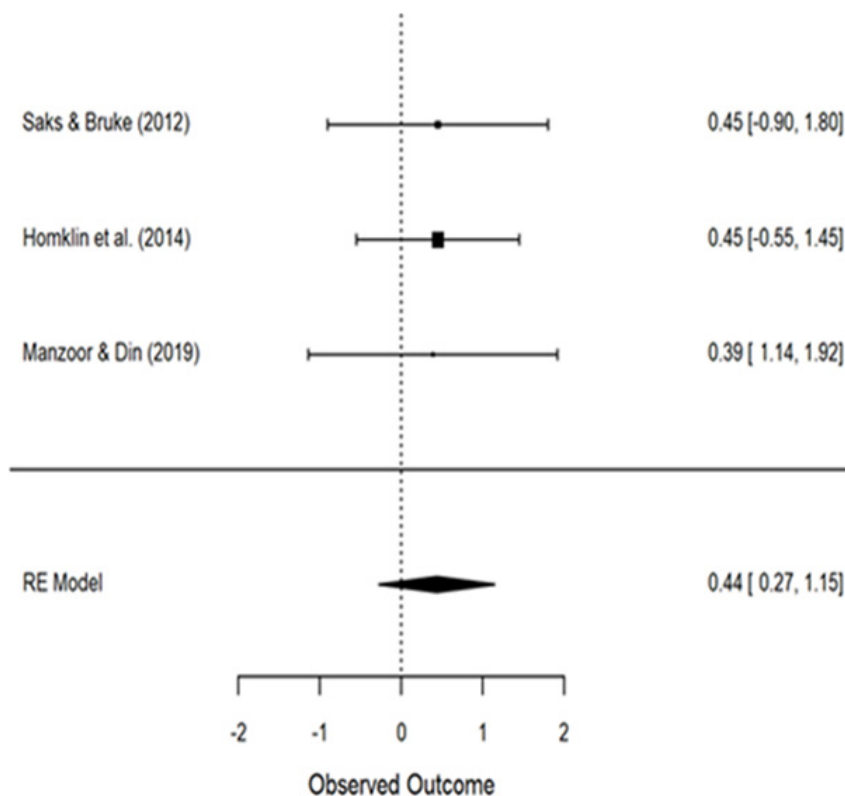


Figure 6: Forest Plot Outcome Behavior to Result

relationships (Alsalamah & Callinan, 2021; Baluku, Matagi & Otto, 2020; Costabella, 2017; Manzoor & Din, 2019). Therefore, the study meta-analytically analyzed the relationships among Kirkpatrick's four levels.

The PRISMA model was used to conduct the systematic review. The researcher fitted the random-effects model using the maximum likelihood estimation. Initially, the relationship between reaction and learning was identified via prior literature. Based on the estimation, about thirty studies confirmed that there exist a positive association exist between two initial levels, i.e., trainee reaction and learning. Secondly, the association between learning and behavior was identified via prior literature. Based on the forest plot estimation, almost twenty six studies confirmed the positive association between trainee learning and behavior. Lastly, association between behavior and result was estimated based on the correlation values of the prior studies. It was discovered that about three studies confirmed a positive correlation between trainee behavior and trainee result. The study revealed majority of studies revealed that a positive association exists between Kirkpatrick four levels of training evaluating model.

5.1 Theoretical Contribution

The authors have reported the several gaps in the current study. These gaps are comprised of, first, mix findings revealed in the literary work pertaining to the linkages among four levels of Kirkpatrick model (Alsalamah & Callinan, 2021; Baluku, Matagi & Otto, 2020; Costabella, 2017; Manzoor & Din, 2019). Second, neglecting to efficacy of measuring the level three, i.e., behavior and level four, i.e. result (Alliger & Janak, 1989; Clement, 1982; Homklin, 2014). Lastly, the existence of little studies pertaining to the Kirkpatrick training evaluating model in the social sciences domain. This research aimed to bridge gaps in the literature by first, taking the last thirty years literary data pertaining to the Kirkpatrick model, i.e. year (1981-2021). Secondly, by representing, incorporating and reporting the literary data via PRISMA model. Thirdly, by evaluating the collected data via forest plot and estimating the correlation values of the factors by random-fixed effect. Additionally, the investigator tried to fill the empirical gaps by connecting Kirkpatrick four levels of training effectiveness. The study mitigates the gaps of the mixed findings pertaining to the linkages among four levels of Kirkpatrick model. This study offers the accommodative literary work regarding Kirkpatrick's four levels of the training effectiveness. Moreover, the study may deliver valuable knowledge of training effectiveness and the baseline for training evaluation to academia and industry.

6. Conclusion

The study's goal was to undertake a systematic review (meta-analysis) of the KTEM. The PRISMA model was used to carry out a systematic review to meet the research goal. A complete sample (n=8825) considered 41 studies (n=41) about the Kirkpatrick model from 1982 to 2021. The Web of Science, Scopus, and Science Direct were selected as repositories for articles pertinent to the study's variables. The researcher fitted the random-effects framework using the maximum likelihood estimation. According to the forest plot findings, most studies have established a positive relationship between trainee reaction and learning. Secondly, many research studies revealed that there is a strong relationship between trainee learning and behavior. Third, a clear correlation between trainee behavior and trainee result was seen in all three studies. The study results significantly impact practical investigations and the HRD practitioners. The study evaluations can help the training programs and suggest the best HRD tactics for advancing and strengthening its trainees. The study findings have provided important data on training efficacy and essential standards for assessing upcoming capacity-building strategies for training. The KTEM model may provide substantial evidence that increases the transparency about training benchmark selection and assessments. The four-level framework of evaluating training may be

used by experienced trainers, decision-makers, and relevant training administrators. To ensure acceptance and successful training transfer, there must be a sufficient level of perceived practical relevance. It is suggested that when undertaking or delivering training courses, the training administrators should take into account and gauge the training effectiveness using the Kirkpatrick framework, which measures trainees' reaction, learning, behavior, and result. To increase the efficiency of training, it is also necessary to carefully assess the standards of social and professional assistance. The research offers the accommodative literary work regarding Kirkpatrick's four levels of the training, evaluating model, including trainees' reaction, trainees' learning, trainees' behavior, and trainees' result.

6.1 Limitations and Future Area

This study has few shortcomings that urge to be highlighted for studies in this area. The meta-analysis was only conducted for the period of (1982-2021) by taking only 41 quantitative based studies. This figure would slightly generalize the research findings because the qualitative based studies are ignored due to the statistical nature of systematic review. In future, the qualitative based study findings may also be incorporated for investigation. Furthermore, researchers must consider employing a mixed-method in the future to gain a thorough grasp of the Kirkpatrick model simultaneously. Additionally, a vast and varied sample in conjunction with sophisticated data processing techniques might increase the possibility that the study's findings will be more generalizable.

References

- Abdelhakim, A. S., Jones, E., Redmond, E. C., Griffith, C. J., & Hewedi, M. (2018). Evaluating cabin crew food safety training using the Kirkpatrick model: an airlines' perspective. *British Food Journal*, 120(7), 1574-1589.
- Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of criteria: Thirty years later. *Personnel Psychology*, 42(4), 331-341.
- Alliger, G. M., Tannenbaum, S. I., Bennett, Jr., W., Traver, H., & Shotland, A. (1997). A meta-analysis on the relations among training criteria. *Personnel Psychology*, 50(4), 341-358.
- Alsalamah, A., & Callinan, C. (2021). Adaptation of Kirkpatrick's four-level model of training criteria to evaluate training programs for head teachers. *Education Sciences*, 11(3), 116.
- Alsalamah, A., & Callinan, C. (2021). The Kirkpatrick model for training evaluation: bibliometric analysis after 60 years (1959-2020). *Industrial and Commercial Training*, 1(2), 36-63.
- Baldwin, T. T. (1992). Effects of alternative modeling strategies on outcomes of interpersonal-skills

- training. *Journal of Applied Psychology*, 77(2), 147.
- Balstad, M. T., & Berg, T. (2020). A long-term bibliometric analysis of journals influencing management accounting and control research. *Journal of Management Control*, 30(4), 357-380.
- Baluku, M. M., Matagi, L., & Otto, K. (2020). Exploring the link between mentoring and intangible outcomes of entrepreneurship: the mediating role of self-efficacy and moderating effects of gender. *Frontiers in Psychology*, 11, 1556.
- Bates, R. A., Holton, E. F. III, Seyler, D. A. & Carvalho, M. A. (2000). The role of interpersonal factors in the application of computer-based training in an industrial setting. *Human Resource Development International*, 3(1), 19-43.
- Bell, B. S., & Ford, J. K. (2007). Reactions to skill assessment: The forgotten factor in explaining motivation to learn. *Human Resource Development Quarterly*, 18(1), 33-62.
- Boet, S., Bould, M.D., Bruppacher, H.R., Desjardins, F., Chandra, D.B., & Naik, V.N. (2011). Looking in the mirror: Self-debriefing versus instructor debriefing for simulated crises. *Critical Care Medicine*, 39, 1377-1381.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97-111.
- Cahapay, M. B. (2021). Kirkpatrick model: Its limitations as used in higher education evaluation. *International Journal of Assessment Tools in Education*, 8(1), 135-144.
- Cannon-Bowers, J. A., & Salas, E. (1997). A framework for developing team performance measures in training. In *Team performance assessment and measurement* (pp. 57-74). Psychology Press.
- Cannon-Bowers, J. A., Salas, E., Tannenbaum, S. I., & Mathieu, J. E. (1995). Toward theoretically based principles of training effectiveness: a model & initial empirical investigation. *Military Psychology*, 7(2), 141-164.
- Chronister, C., & Brown, D. (2012). Comparison of simulation debriefing methods. *Clinical Simulation in Nursing*, 8(2), 69-81.
- Clement, R. W. (1982). Testing the hierarchy theory of training evaluation: An expanded role for trainee reactions. *Public Personnel Management*, 11(2), 176-184.
- Costabell, L. M. (2017). Do high school graduates benefit from intensive vocational training?. *International Journal of Manpower*, 4(3), 55-62.
- Dachner, A. M., Ellingson, J. E., Noe, R. A., & Saxton, B. M. (2021). The future of employee development. *Human Resource Management Review*, 31(2), 100732.
- Deeks, J. J. (2002). Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine*, 21(11), 1575-1600.

- Dreifuerst, K.T. (2012). Using debriefing for meaningful learning to foster development of clinical reasoning in simulation. *Journal of Nursing Education*, 51, 326-333.
- Fisher, S. L., & Ford, J. K. (1998). Differential effects of learning effort & goal orientation on two learning outcomes. *Personnel Psychology*, 51, 397-420.
- Fisher, S. L., Wasserman, M. E., & Orvis, K. A. (2010). Trainee reactions to learner control: an important link in the e-learning equation. *International Journal of Training and Development*, 14(3), 198-208.
- Frayne, C. A., & Geringer, J. M. (2000). Self-management training for improving job performance: A field experiment involving salespeople. *Journal of Applied Psychology*, 85(3), 361.
- Grant, J.S., Dawkins, D., Molhook, L., Keltner, N.L., & Vance, D.E. (2014). Comparing the effectiveness of video-assisted oral debriefing and oral debriefing alone on behaviors by undergraduate nursing students during high-fidelity simulation. *Nurse Education in Practice*, 14, 479-484.
- Grant, J.S., Moss, J., Epps, C., & Watts, P. (2010). Using video-facilitated feedback to improve student performance following high-fidelity simulation. *Clinical Simulation in Nursing*, 6(2), 112-131.
- Gully, S. M., Payne, S. C., Koles, K. L. K., & Whiteman, J. A. K. (2002). The impact of error training and individual differences on training outcomes: An attribute-treatment interaction perspective. *Journal of Applied Psychology*, 87(1), 143-155.
- Hazan-Liran, B., & Miller, P. (2020). The relationship between psychological capital and academic adjustment among students with learning disabilities and attention deficit hyperactivity disorder. *European Journal of Special Needs Education*, 1-14.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic press. New York.
- Ho, A. D., Arendt, S. W., Zheng, T., & Hanisch, K. A. (2016). Exploration of hotel managers' training evaluation practices and perceptions utilizing Kirkpatrick's and Phillips's models. *Journal of Human Resources in Hospitality & Tourism*, 15(2), 184-208.
- Hoffmann, T. C., Oxman, A. D., Ioannidis, J. P., Moher, D., Lasserson, T. J., Tovey, D. I., ... & Glasziou, P. (2017). Enhancing the usability of systematic reviews by improving the consideration and description of interventions. *Business Management Journal*, 35(8), 1-7.
- Holton, E. F. (1996). The flawed four level evaluation model. *Human Resource Development Quarterly*, 7(1), 5-21.
- Homklin, T. (2014). Training effectiveness of skill certification system: The case of automotive industry in Thailand. *Unpublished doctoral dissertation*. Hiroshima University, Hiroshima, Japan.
- Lalkhen, A. G., & McCluskey, A. (2008). Clinical tests: sensitivity and specificity. *Continuing education in anaesthesia critical care & pain*, 8(6), 221-223.
- 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*

- (IJMESS), 8(4), 266-279.
- Liao, W.C. & Tai, W. T. (2006). Work-related justice, motivation to learn, & training outcomes. *Social Behavior & Personality*, 34(5), 545-556.
- Liebermann, S. & Hoffmann, S. (2008). The impact of practical relevance on training transfer: evidence from a service quality training program for German bank clerks. *International Journal of Training & Development*, 12(2), 74-86.
- Lim, H., Lee, S. G., & Nam, K. (2007). Validating E-learning factors affecting training effectiveness. *International Journal of Information Management*, 27(1), 22-35.
- Lin, Y. T., Chen, S. C., & Chuang, H. T. (2011). The effect of organizational commitment on employee reactions to educational training: An evaluation using the Kirkpatrick four-level model. *International Journal of Management*, 28(3), 926.
- Manzoor, S. R. & Din, Z.U (2019). Measuring the training effectiveness in the police sector of Pakistan: A Kirkpatrick model intervention. *Journal of Managerial Sciences*, 13(2). 30-45.
- Mariani, B., Cantrell, M. A., Meakim, C., Prieto, P., & Dreifuerst, K. T. (2013). Structured debriefing and students' clinical judgment abilities in simulation. *Clinical Simulation in Nursing*, 9(5), e147-e155.
- McEvoy, G. M. (1997). Organizational change and outdoor management education. *Human Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management*, 36(2), 235-250.
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1), 1-9.
- Moral Muñoz, J. A., Herrera Viedma, E., Santisteban Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: *An up-to-date Review*. 29(1), 118-140.
- Orvis, K. A., Fisher, S. L., & Wasserman, M. E. (2009). Power to the people: using learner control to improve trainee reactions & learning in web-based instructional environments. *Journal of Applied Psychology*, 94(4), 960-971.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic reviews*, 10(1), 1-11.
- Patki, S., Sankhe, V., Jawwad, M., & Mulla, N. (2021). Personalised Employee Training. In *2021 International Conference on Communication information and Computing Technology (ICCICT)* (pp. 1-6). IEEE.
- Reed, S.J. (2015). Written debriefing: Evaluating the impact of the addition of a written component when debriefing simulations. *Nurse Education in Practice*, 15, 543-548.
- Reed, S.J., Andrews, C.M., & Ravert, P. (2013). Debriefing simulations: Comparison of debriefing with

- video and debriefing alone. *Clinical Simulation in Nursing*, 9, 88-103.
- Richman-Hisrich, W. (2001). Post training interventions to enhance transfer: the moderating effects of work environments, *Human Resource Development Quarterly*, 12(2), 105-120.
- Saks, A. M., & Burke, L. A. (2012). An investigation into the relationship between training evaluation and the transfer of training. *International Journal of Training and Development*, 16(2), 118-127.
- Sampson, M., Tetzlaff, J., & Urquhart, C. (2011). Precision of healthcare systematic review searches in a cross-sectional sample. *Research Synthesis Methods*, 2(2), 119-125.
- Samwel, J. O. (2018). Impact of employee training on organizational performance—case study of drilling companies in geita, shinyanga and mara regions in tanzania. *International Journal of Managerial Studies and Research*, 6(1), 36-41.
- Savoldelli, G.L., Naik, V.N., Park, J., Joo, H.S., Chow, R., & Hamstra, S.J. (2006). Value of debriefing during simulated crisis management: Oral versus video-assisted oral feedback. *Anesthesiology*, 105, 279-285.
- Schulze, R. (2004). *Meta-analysis-A comparison of approaches*. Hogrefe Publishing.
- Shareefa, M., & Moosa, V. (2020). The Most-Cited Educational Research Publications on Differentiated Instruction: A Bibliometric Analysis. *European Journal of Educational Research*, 9(1), 331-349.
- Shinnick, M.A., Woo, M., Horwich, T.B., & Steadman, R. (2011). Debriefing: The most important component in simulation? *Clinical Simulation in Nursing*, 7, 77-90.
- Sitzmann, T., Bell, B. S., Kraiger, K., & Kanar, A. M. (2009). A multilevel analysis of the effect of prompting self-regulation in technology-delivered instruction. *Personnel Psychology*, 62(4), 697-734.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological bulletin*, 87(2), 245.
- Sulsky, L. M., & Kline, T. J. B. (2007). Understanding frame-of-reference training success: a social learning theory perspective. *International Journal of Training & Development*, 11(2), 121-131.
- Tan, J. A., Hall, R. J., & Boyce, C. (2003). The role of employee reactions in predicting training effectiveness. outcomes. *Human Resource Development Quarterly*, 14(4), 397-411.
- Tracey, J. B., Hinkin, T. R., Tannenbaum, S., & Mathieu, J. E. (2001). The influence of individual characteristics & the work environment on varying levels of training outcomes. *Human Resource Development Quarterly*, 12(1), 5-23.
- Van Heukelom, J.N., Begaz, T., & Treat, R. (2010). Comparison of postsimulation debriefing versus in-simulation debriefing in medical simulation. *Simulation in Healthcare*, 5(2), 91-97.
- Velada, R., & Caetano, A. (2007). Training transfer: the mediating role of perception of learning. *Journal of European Industrial Training*. 31(4), 283-296.
- Warr, P., & Bunce, D. (1995). Trainee characteristics and the outcomes of open learning. *Personnel*

Psychology, 48(2), 347-375.

- Warr, P., Allan, C., & Birdi, K. (1999). Predicting three levels of training outcome. *Journal of Occupational and Organizational Psychology*, 72(3), 351-375.
- Weaver, A. (2015). The effect of a model demonstration during debriefing on students' clinical judgment, self-confidence, and satisfaction during a simulated learning experience. *Clinical Simulation in Nursing*, 11(1), 20-26.
- Welke, T.M., LeBlanc, V.R., Savoldelli, G.L., Joo, H.S., Chandra, D.B., Crabtree, N.A., & Naik, V.N. (2009). Personalized oral debriefing versus standardized multimedia instruction after patient crisis simulation. *Anesthesia and Analgesia*, 109, 183-189.
- Wexley, K. N., & Baldwin, T. T. (1986). Post training strategies for facilitating positive transfer: An empirical exploration. *Academy of Management Journal*, 29(3), 503-520.
- Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology*, 75(3), 315.
- Zielińska-Tomczak, Ł., Przymuszała, P., Tomczak, S., Krzyśko-Pieczka, I., Marciniak, R., & Cerbin-Koczowska, M. (2021). How do dieticians on instagram teach? The potential of the Kirkpatrick model in the evaluation of the effectiveness of nutritional education in social media. *Nutrients*, 13(6), 2005-2022.

