JCSHD
COGNITIVE SCIENCES AND HUMAN DEVELOPMENT

A Critical Review on Training Evaluation Models: A Search for Future Agenda

Badriya Ambu-Saidi*, Fung, C.Y., Turner, K., & Lim, A.S.S.
Faculty of Business, Design and Arts, Swinburne University of Technology, Australia.

ABSTRACT

This paper delves into the intricate landscape of training transfer evaluation, exploring some models, strengths, and areas for improvement in assessing the effectiveness of training programs. This paper aims to analyse and compare training transfer evaluation models using a scoping review methodology. By examining their strengths, weaknesses and unique characteristics, the paper seeks to pinpoint opportunities for enhancement. The discussion centres around seminal works such as Kirkpatrick's Four-Level Model, Kaufman and Keller's Five Levels of Evaluation, the Phillips Return on Investment Model, Warr et al.'s Context, Input, Reaction and Outcome Model and Brinkerhoff's Six-Stage Model, Bushnell's input, process, output model. The paper underscores the importance of a comprehensive and adaptable approach to training transfer evaluation, emphasising the need for hybrid models that integrate strengths while addressing weaknesses. The exploration extends to measuring tools and research methods that enhance evaluation practices. The sequential explanatory mixed methods design emerges as an exemplar of a research methodology that seamlessly combines quantitative and qualitative approaches to offer a richer understanding of training transfer. As the paper concludes, it advocates for continuous research efforts to refine models, incorporate emerging technologies, and align evaluation practices with learning and organisational development dynamics. By revealing gaps in current knowledge and identifying previously unknown areas for improvement in training transfer, this paper contributes novel insights to the field of training transfer.

Keywords: teacher training, training transfer, training evaluation models, sequential explanatory mixed methods design

ARTICLE INFO
Email address: badriya.ambusaidi@moe.om (Badriya Ambu-Saidi)
*Corresponding author
https://doi.org/10.33736/jcshd.6336.2024
e-ISSN: 2550-1623

Manuscript received: 10 December 2023; Accepted: 27 March 2024; Date of publication: 31 March 2024

1 INTRODUCTION

Training is essential in human resource practices since it is a tool for personal development (Bednall & Sanders, 2017; Choi, 2017) and organisational performance (Singh et al., 2017). Employees can build and cultivate their knowledge and skills through training, thus improving their productivity (Riley et al., 2017). To keep pace with change, Chen, J. & Hou J. (2021) emphasised that teachers in the educational field must acquire and cultivate a wide range of skills essential for their success and competitiveness in their profession. In addition, teachers must be well-prepared considering uncertainties and challenges in the work environment resulting from global competition and organisational restructuring (i.e., merging and downsizing). They thus may need to equip themselves with higher-order skills (e.g., leadership and psychological skills). Burke and Hutchins (2008) reported that approximately 40% of trainees need to immediately transfer their newly acquired skills upon completing a training program.

Furthermore, within a year, about 70% of trainees discontinue using the skills they learned, and, on average, only half of the training investment yields performance improvement (Al et al., 2018). Elnaga and Imran (2013) found that merely 10% of training investments led to observable behavioural change. These findings underscore a persistent need for return on training investment. Consequently, scholars and organisational decision-makers have directed considerable attention towards training transfer (Al Jabri et al., 2018).

In the context of this paper, transfer of training (TOT) is defined as the application, generalisation and maintenance of newly learned knowledge and skills on the job over a lengthy period (Baldwin & Ford, 2017). Despite these positive aims, teacher professional development programs may fail to improve teachers' performance if their content is of mediocre quality or limited relevance, if they are delivered ineffectively, if they lack follow-up to help teachers translate learning into practice or if the programs fail to hold trainees accountable for their teaching (Kim et al., 2020). In addition, training has gained a lot of attention and investment. Studies have indicated that educational organisations in the United States of America spent around USD 1,765 per in-service teacher (USD 182 billion or 2.38 % of total payroll) on teacher training and development in 2016 (American Society for Training & Development, 2017). Furthermore, a 2016 study conducted on behalf of the United Kingdom Commission for Employment and Skills (UKCES, 2016) declared that educational organisations in the UK invested nearly £63 billion in professional development programs for teachers. However, teachers have difficulty applying what they have learned in the classroom. (Blume, Ford, Baldwin & Huang, 2010; Noe, 2016). Employees, including teachers, can experience challenges with training transfer when they attend training programs (Holton, 2005). TOT and its evaluation are a lengthy process, and human resources (HR) and human resources development (HRD) practitioners take plenty of time and effort to handle them. HRD experts are expected to plan, deliver, and evaluate learning interventions to link training and improved performance. Limited evidence exists to prove the success of training transfer (Al-Omairi, 2021; Chen & Hou, 2021; and Roca-Puig & Llorens-Montes, 2021).

Transfer of training is seen as a continuing problem in educational organisations as policymakers and stakeholders are concerned about this ongoing issue (Antoniou & Kyriakides, 2013). Teachers often encounter challenges with training transfer when they participate in training programs

(Holton, 2005). Although they acquire valuable knowledge and skills during the training, they frequently struggle to effectively practice, apply, and fully master what they have learned once they return to their work environment (Al-Hakamani, 2011; Al et al., 2018). Consequently, a substantial body of research exists exploring the factors that influence the transfer of training among teachers. Much of this research has centred on training design and delivery considerations (Baldwin & Ford, 2010; Malik & Grover, 2014; Hafeez & Akbar, 2015). These studies investigate several aspects of training programs, such as their structure, content, and instructional methods, to identify factors that enhance or hinder the transfer of learned knowledge and skills into practical teaching contexts. By examining training design and delivery issues, researchers aim to provide insights and recommendations to optimise training programs and promote successful transfer of training among teachers (Noe, 2016). Therefore, attention should be directed to factors beyond the training program, such as trainee characteristics and work environment (Noe, 2016).

This paper aims to analyse and compare key training transfer evaluation models, highlighting their strengths, weaknesses, and distinct attributes to identify areas for improvement. It also proposes developing an evaluation model that offers a comprehensive approach to assessing the effectiveness of training programs.

2 LITERATURE REVIEW

2.1 Teacher Training

The literature on teacher training highlights the importance of training in improving organisational performance (Pak et al., 2016) and teaching (Christophersen et al., 2016; Dolev & Leshem, 2017; Gavish, 2017). That is, through training, teachers improve their teaching practice. Al-Aufi (2014) states that training programs can help new teachers develop professionally, including enhancing teachers' knowledge, attitudes, beliefs, and practices.

According to Dolev and Leshem (2017), practical teacher training is a cornerstone of educational excellence, particularly for new teachers' careers. This training equips them with the essential tools to navigate the complexities of the classroom. By imparting pedagogical insights and proven teaching strategies, training empowers new educators to deliver content effectively and cultivate engaging and dynamic learning environments (Dolev & Leshem, 2017). In line with this, Gavish (2017) stated that classroom management skills, differentiation techniques, and inclusion practices are instilled during training programs, enabling teachers to meet the diverse needs of their students. Furthermore, assessment expertise and feedback proficiency are honed, enabling accurate evaluation of student progress and the delivery of constructive guidance (Gavish, 2017). Through reflective practices encouraged by training, new teachers evolve professionally, staying aligned with research-based methodologies and gaining the confidence needed to tackle challenges (Al-Aufi, 2014). Mentorship, often embedded in training programs, provides invaluable guidance, and fosters competence and compassion in teaching. However, despite the comprehensive training provided, transferring acquired skills into the classroom often needs help. Many educators need help implementing newly learned strategies effectively, indicating a gap between training and application.

Moreover, the evaluation of training transfer remains a concern, with methods for assessing the practical integration of training outcomes needing refinement. Moreover, the effectiveness of introductory training programs for teachers in Arab educational organisations remains uncertain due to inadequate evaluation tools and a lack of coaching and follow-up in the training transfer process (Al-Jardani, 2015 & Al-Ghatrifi, 2016). Training programs in some Arab educational organisations are often not acknowledged as vital organisational functions that contribute to the organisation's overall success (Altarawneh & Aseery, 2016). In Oman, teacher training and professional development programs encounter challenges, including a need for more investigation into training transfer and insufficient organisational and social support (Al-Balushi, 2017 & Al-Jardani, 2015). Recognising and addressing these challenges is essential to ensure that teacher training has meaningful and lasting impacts on classroom instruction and student learning outcomes.

2.2 Transfer of Training

Transfer of training has been well defined in the literature. Baldwin and Ford (2010) defined transfer as the generalisation of the skills learned during a learning event to the workplace and the maintenance of the acquired skills over time. Similarly, Avalos (2011) defined transfer of training as the successful and ongoing application of training knowledge and skills to a trainee's job. This means that training transfer refers to trainees' practical and continuous application of what they learned in training, including knowledge and skills, to the jobs (Al-Omairi, 2021 & Noe, 2016). Another definition by Barnett (2005) describes the transfer as the knowledge used after training is complete, and employees should not only practice appropriate behaviour in training but also demonstrate it in the workplace. Also, according to Grossman and Salas (2011), the transfer is the process of implementing knowledge, skills, attitudes, and other behaviours acquired in the training program into the workplace. Thus, based on those definitions, one can conclude that training transfer entails application, maintenance of newly acquired knowledge and skills, and generalizability. (Coetsee & Eiselen, 2006).

The continuity of applying new skills is vital for learning and transferring such skills (Noe, 2016). Therefore, more is needed to learn or apply a new skill, but employees should also maintain the learning behaviour. Learning behaviour resembles long-term anticipation and commitment to maintain and utilise skills (Al-Aufi, 2012). Moreover, based on Alfonso and Ramirez's (2021) view, learning, or mastering the information covered during a training program, using the new skills and knowledge on the job, and maintaining the change in behaviour over time are the three steps involved in transfer. Although most definitions focus on the actual transfer from training to work, training can only be transferred if learning occurs. Any transfer benefits or gains can only be realised if the transfer is maintained over time. Hence, we can better understand transfer as a dynamic and complex process (Bendall & Sanders, 2017). The coming section sheds light on some evaluation models used to evaluate training transfer. It proposes using a sequential explanatory mixed methods design to improve the evaluation of the transfer of training for teachers.

2.3 Training Transfer Evaluation Models

2.3.1 Kirkpatrick's Four-Level Model

Kirkpatrick's Four-Level Model is a widely used framework for evaluating the effectiveness of training and learning programs. Developed by Donald L. Kirkpatrick in the 1950s, the model provides a structured approach to assess the impact of training initiatives on learners and their organisations (Milne, 2007). The model consists of four levels, each representing a different aspect of evaluation:

1. Level 1: Reaction: This level focuses on participants' immediate reactions and perceptions of the training. It involves gathering feedback on the training's content, delivery, materials, and overall experience. This feedback can be collected through surveys, questionnaires, and discussions. Evaluating participants' reactions helps trainers and organisations understand how engaging and relevant the training is to learners and whether any adjustments are needed.

2. Level 2: Learning: Level 2 assesses how participants have acquired the intended knowledge, skills, and attitudes from the training. This involves measuring learning outcomes through tests, assessments, skill demonstrations, and observations. The goal is to determine whether the training has effectively transferred current information and abilities to the participants.

3. Level 3: Behaviour: This level shifts the focus to the workplace environment. It examines whether participants apply the acquired knowledge and skills on the job. Evaluating behaviour change often requires ongoing observations, surveys of supervisors and peers, and other methods to gauge the extent of the practical application of the training content.

4. Level 4: Results: The highest level of evaluation, Level 4, looks at the broader impact of training on the organisation's goals and objectives. This could include metrics such as increased productivity, improved customer satisfaction, reduced error rates, or enhanced overall performance. The goal is to determine the tangible benefits the training has brought to the organisation and whether the investment in training has been worthwhile.

Kirkpatrick's model is hierarchical, with each level building on the previous one. It emphasises the importance of evaluating training beyond participants' reactions and seeks to measure the impact on performance and organisational outcomes (Saks & Burke, 2012). Training and development professionals have widely adopted the model to guide training programs' design, implementation, and evaluation. However, it is essential to note that while the model provides a valuable framework, its application might vary based on the specific context and goals of the training initiative (Nickols, 2005; Khalid et al., 2012; Saks & Burke, 2012). Kirkpatrick's Four-Level Model of training evaluation has been widely utilised in assessing the effectiveness of teachers' training programs (Kirkpatrick, 1994). Scholars and practitioners have applied this model to evaluate teachers' professional development, from initial training programs to ongoing

professional learning initiatives—the model's four levels are reaction, learning, behaviour, and results. The model has provided a structured framework for evaluating the impact of teachers' training interventions. At the reaction level, feedback from teachers regarding their satisfaction with training content, delivery methods, and overall experience has been collected and analysed (Saks & Burke, 2012). This feedback has helped training providers tailor their programs to meet teachers' needs and preferences better.

Moving to the learning level, assessments of teachers' knowledge acquisition, skill development, and competency enhancement have been conducted (Khalid et al., 2012). These assessments have allowed evaluators to gauge the effectiveness of training in enhancing teachers' understanding of pedagogical concepts, subject matter knowledge, and instructional strategies.

At the behaviour level, observations of teachers' classroom practices, instructional techniques, and implementation of newly acquired skills have been undertaken (Joyce & Showers, 2002). Evaluators have examined whether teachers effectively translate their learning from training sessions into their day-to-day teaching practices and whether there are observable improvements in student engagement, learning outcomes, and classroom management.

Finally, at the results level, the impact of teachers' training on broader educational outcomes, such as student achievement, school performance, and teacher retention rates, has been assessed (Hattie, 2009). By examining these outcomes, stakeholders have determined teachers' training programs' overall effectiveness and return on investment.

Criticism of Kirkpatrick's four-level model

The model has faced critique on multiple fronts, encompassing its inherent incompleteness, the presumption of causality, and the postulation of escalating information significance as Kirkpatrick's four-tiered framework for training evaluation is traversed (Bates, 2004). Scholars such as Bates (2007) and Guerci et al. (2010) contend that Kirkpatrick's model offers an oversimplified perspective on training effectiveness, inadequately capturing the intricate process of training evaluation. They further assert its failure to incorporate the influence of individual and organisational factors on training evaluation. Despite these objections (Bates, 2004; Holton, 1996; Hung, 2010), it is noteworthy that this model holds pre-eminence among academics (Phillips, 1996) and finds widespread adoption in organisational contexts (Bates, 2004).

Regarding its incompleteness, the model neglects the impact of individual and contextual factors on training efficacy (Bates, 2004). The effectiveness of training is a confluence of organisational, individual, and training design and delivery elements before, during, and post training (Tannenbaum & Yukl, 1992; Cannon-Bowers et al., 1995; Ford & Kraiger, 1995; Salas & Cannon-Bowers, 2001). This study's training design and delivery aspects are designated "training characteristics." However, the model assumes that exploring these factors is dispensable when employing Kirkpatrick's model for evaluating training effectiveness (Bates, 2004).

Furthermore, the model overlooks the influence of the intention to transfer learning, which guides learner conduct in the workplace. Clemenz (2001) proposes that this intention serves as the bridge between level 1 ("reaction") and level 3 ("behaviour") within Kirkpatrick's model. Foxon (1993) posits that the transfer of training process originates from the learner's intention to apply learning. Intention to transfer learning is delineated as "the trainees' deliberate intent to engage in specific conduct facilitating the application of their acquired skills" (Bansal & Thakur, 2013, p. 56). As per Ajzen and Fishbein (1977), intent serves as a direct precursor to action or behaviour. Ajzen (1991) further elucidates that intent precedes behaviour. Consequently, trainees are inclined to formulate an intent to behave in a particular manner post-training completion (Yamkovenko & Holton, 2010).

Additionally, Kirkpatrick (1994) asserts a causal linkage among the four levels, positing that favourable reactions lead to learning, learning results in sought-after behavioural shifts in the professional setting, and so forth (Bates, 2004; Kirkpatrick & Kirkpatrick, 2006; Smith & Johnson, 2023). Thus, the lower strata of Kirkpatrick's model necessitate preliminary evaluation to glean meaningful outcomes from the assessment (Alliger & Janak, 1989). Reaction evaluation takes precedence, as positive reactions towards a training initiative might motivate future participation, while unfavourable reactions could deter learners from engagement or completion (Reio et al., 2017). This implies that a robust evaluation of the higher echelons of the model necessitates a prior assessment of the foundational levels. Each of the four evaluation outcomes yields distinct insights into training, varying in utility depending on the evaluation's purpose (Stewart & Brown, 2011).

Thirdly, according to Alliger and Janak (1989), Kirkpatrick's model presupposes that subsequent levels furnish more valuable insights into the training program than their antecedent counterparts. These four tiers are arranged in ascending order, yielding a hierarchical model (Reio et al., 2017). It is, thus, erroneous to bypass Level 1 (reaction) and Level 2 (learning) and exclusively appraise Level 3 (behaviour) and Level 4 (results) (Kirkpatrick & Kirkpatrick, 2006). The proposition of escalating knowledge significance serves as one of the bases for criticism of Kirkpatrick's model (Bates, 2007). Nonetheless, empirical research lacks substantial evidence to substantiate the assumption that each successive level confers superior data than its preceding counterpart (Bates, 2004).

While the critique holds merit, viewing it in a broader context is essential. Rather than dismissing the model outright for its perceived incompleteness, one can argue that individual and contextual factors should be considered in conjunction with the model rather than in isolation. Kirkpatrick's Four-Level Model provides a structured and hierarchical framework for evaluating training impact, and incorporating additional factors can enhance its applicability. For instance, individual characteristics, such as motivation, prior knowledge, and learning preferences, can be integrated into the evaluation process. Similarly, contextual factors, including the organisational culture, support systems, and job requirements, can be considered alongside the model's four levels. This comprehensive approach recognises the complexity of training effectiveness and ensures a more comprehensive assessment that aligns with the diverse dynamics of the learning environment.

Many alternative models have been conceptualised and formulated in response to the criticisms directed towards the Kirkpatrick model. This array of models is documented in the works of

Brinkerhoff (1987, 2003), Bushnell (1990), Hamblin (1974), Holton (1996), Kaufman et al. (1996), Kraiger et al. (1993), Phillips (1996), Stufflebeam (1983), and Warr et al. (1970). Subsequent sections will delve into a comprehensive investigation of these models above.

2.3.2 Kaufman and Keller's Five Levels of Evaluation

Kirkpatrick's four-level framework has been criticised for omitting the societal impact of training (Stokking, 1996). This model has faced criticism for its perceived incompleteness, focusing narrowly on training evaluation (Watkins et al., 1998). An extended rendition proposed by Kaufman and Keller (1994) introduces a broader perspective that encompasses societal value addition and continuous enhancement, deviating from summative measurement (Watkins et al., 1998). This extended model supplements Kirkpatrick's four tiers by expanding the reaction level scope to encompass enabling and reaction (input-process) facets. A fifth level is also introduced to gauge societal outcomes and the training's influence on society (Russ-Eft et al., 1997). Kaufman et al. (1996) assert that applying evaluation levels beyond training opens avenues for considering other performance-enhancing interventions. Thus, this model considers both internal and external consequences of training, linking them to performance and organisational advancement (Passmore & Velez, 2012). Kaufman and Keller (1994) contend that Kirkpatrick's original model understates training's societal impact, consequently devaluing resources, and methodologies.

According to Russ-Eft et al. (1997) and Kaufman et al. (1996), evaluating societal outcomes and expanding the reaction level in Kirkpatrick's model are instrumental in determining trainee contentment with training resources and methodologies. It also assesses whether beneficial training outcomes are realised and whether the contributions of training to society hold merit.

In the context of teachers' training, Kaufman and Keller's model has been utilised to assess the internal aspects of training and its broader societal outcomes (Russ-Eft et al., 1997). By incorporating a fifth level to gauge societal outcomes and the training's influence on society, this model acknowledges the broader impact of teachers' training beyond individual learning outcomes (Kaufman et al., 1996). Furthermore, evaluating societal outcomes and expanding the reaction level in Kaufman and Keller's model is instrumental in determining teachers' satisfaction with training resources and methodologies and assessing whether beneficial training outcomes are realised (Russ-Eft et al., 1997).

Kaufman's model of evaluation consists of five levels: (1) enabling and reaction, also termed input and process, (2) acquisition, (3) application, (4) organisational output, and (5) societal outcomes. At Level 1, the reaction concept is broadened to encompass enabling and reaction (input and process). This level introduces a distinction between input concerning the quality of available organisational resources and the acceptability and efficiency of methods and resources employed in a process. This distinction provides evaluators, managers, and decision-makers with essential insights to enhance organisational training and educational efforts iteratively. Level 2 pertains to "acquisition," measuring the acquisition of competencies. Level 3 pertains to the application of learned skills within the workplace, while Level 4 evaluates organisational output, quantifying the contributions or yield of the organisation. Finally, Level 5 evaluates societal matters,

encompassing responsiveness to society and clients, consequences, and returns (Jamjoom & Al-Mudimigh, 2011; Kaufman et al., 1996).

However, Stokking (1996) argues that certain aspects within the expanded Kirkpatrick model need to be revised and need more clarity, suggesting further refinement. Disparities arise between desired chronological activities and levels' facets, leading to ambiguity, particularly regarding training implementation conditions. This model incorporates implementation and learning objectives and their accomplishments into Level 2 (acquisition) due to their role as indicators of training effectiveness and appropriate course deployment (Stokking, 1996). Moreover, Kaufman's model carries theoretical underpinnings but offers limited practical applicability (Topno, 2012). Consequently, this model furnishes information comparable to Kirkpatrick's four-level framework, as contextual factors need to be considered. While Kaufman's model aims to broaden the evaluation scope to include societal impact, the critique raises valid concerns about the need for clarity and refinement. Acknowledging that any model, no matter how comprehensive, may encounter challenges in its application is crucial. Instead of dismissing the model outright, one can argue for a more nuanced interpretation and potential refinements to address these concerns.

2.3.3 The Phillips Return on Investment Model

Kirkpatrick's four-level model must improve its capacity to gauge a training program's economic value and advantages effectively. In response to this limitation, Phillips (1996) introduced a supplementary dimension termed "return on investment" (ROI) as a fifth level to the existing Kirkpatrick framework. Moreover, Phillips expanded the scope of Level 1 to encompass the intentions of trainees to apply acquired knowledge from the training program within their professional contexts. Phillips defines return on investment as "a ratio that juxtaposes the monetary gains with the training costs" (McKenna & Beech, 2014, p. 377).

Phillips (2005) contends that including the ROI level yields valuable data and substantial empirical evidence that substantiates the recouping of expenditures linked to training efforts. This is achieved by demonstrating the financial advantages of training interventions through meticulous cost-benefit analyses (Lockwood, 2001; Chang, 2010). However, it is essential to acknowledge the challenges associated with assessing return on investment in teachers' training programs. Russ-Eft and Preskill (2005) highlight the multifaceted and intricate nature of delineating ROI within complex educational systems. Determining ROI involves a greater degree of subjectivity than objective quantification, and outcomes may not constantly precisely evaluate training investments due to the entanglement of net training benefits with other organisational variables (Wang & Wilcox, 2006). The computation of return on investment is characterised by a greater degree of subjectivity than objective quantification and has yielded outcomes that could be more precise in evaluating training investments. This is primarily due to the entanglement of net training benefits with other variables inherent to organisational systems, rendering their distinct separation challenging, despite the relative ease of determining overall training costs (Wang & Wilcox, 2006).

2.3.4 Warr et al.’ s Context, Input, Reaction and Outcome Model

Kirkpatrick's four-tier model's inadequacy in capturing contextual factors and pre-training inputs prompts Warr et al. (1970) to introduce the Context, Input, Reaction, and Outcome (CIRO) framework for assessing managerial training programs. The CIRO model entails an exhaustive analysis of contextual considerations and potential inputs before evaluating participant reactions (Brewer, 2007; Tamkin et al., 2002). The contextual stage involves an examination of the prevailing operational landscape to discern training requisites and objectives. The input phase involves information regarding feasible training methodologies or techniques, aiding in the optimal selection of training interventions (Brewer, 2007).

In the context of teachers' training, the CIRO model entails an exhaustive analysis of contextual factors and potential inputs before evaluating participant reactions (Brewer, 2007; Tamkin et al., 2002). The contextual stage involves examining the prevailing educational landscape to discern training requisites and objectives specific to teachers' professional development needs. This adaptation ensures that the training program aligns with the unique challenges and goals of the teaching profession.

Comparable to Kirkpatrick's model, the reaction stage in the CIRO framework aligns with participant responses but places greater significance on soliciting recommendations to enhance the training program's responsiveness to participant insights (Brewer, 2007). Like Kirkpatrick's learning, behaviour, and results levels, the outcome stage scrutinises training outcomes encompassing immediate, intermediate, and ultimate impacts (Brewer, 2007; Phillips, 2003; Tamkin et al., 2002).

Nevertheless, while underscoring objectives and resource availability, the CIRO model omits guidance on executing these evaluations (Tzeng et al., 2007). Administered before and after training interventions (Tennant et al., 2002), the model's outcome evaluation may not always be obligatory, a viewpoint advocated by Warr et al. (1970). In addition, it is essential to note that, the study emphasises the comprehensive nature of the CIRO model. While it successfully underscores the importance of considering objectives and resource availability, there needs to be more guidance on evaluating evaluations, as highlighted by Tzeng et al. (2007). This omission raises practical challenges in implementing the model effectively. According to Tennant et al. (2002), the CIRO model recommends evaluation before and after training interventions. However, it acknowledges that the outcome evaluation may sometimes be obligatory, aligning with Warr et al.'s (1970) viewpoint. However, prevailing trends in the human resources development domain favour assessment at the results level (Chang, 2010). For optimal efficacy, the CIRO model necessitates further refinement and development.

2.3.5 Brinkerhoff’s Six-Stage Model

The Kirkpatrick model has faced criticism due to its need for more consideration in assessing training requirements prior to the commencement of training and integrating these assessments into the training design process. As a result, Brinkerhoff (1987) formulated a six-stage evaluation model encompassing the following stages: 1) goal setting, 2) program design, 3) program implementation, 4) immediate outcomes, 5) intermediate or usage outcomes, and 6) impacts and

value assessment. Brinkerhoff's model shares similarities with Kirkpatrick's framework (Bomberger, 2003; Phillips, 2003). However, it extends Kirkpatrick's model by introducing two preliminary phases to conduct a formative assessment of training needs and design (Holton & Naquin, 2005).

In the context of teachers' training, Brinkerhoff's model encompasses six stages: goal setting, program design, program implementation, immediate outcomes, intermediate or usage outcomes, and impacts and value assessment (Holton & Naquin, 2005). The initial stages of goal setting and program design are crucial in aligning the training program with the specific needs and objectives of teachers' professional development.

In Brinkerhoff's model, stage 3 corresponds to Kirkpatrick's "reaction" level, while stage 4 aligns with the "learning" level (Change, 2010). Stage five evaluates the successful application of learned skills in the workplace, while stage six gauges the program's organisational value (Change, 2010; Kumpikaite, 2007). These stages are analogous to Kirkpatrick's levels 3 ("behaviour") and 4 ("results") in the four-level model. Notably, these six stages are interconnected, operating sequentially, where each subsequent stage builds upon the completion of the previous one.

However, it is essential to note that Brinkerhoff's model is most suitable for scenarios where close collaboration between employers and training organisers is feasible. This is due to its pre- and post-training evaluation requirements (Passmore & Velez, 2012). Additionally, the model is well-suited for instances where the evaluation design is integrated into the training process, as stages 1 (goal setting) and 2 (program design) are intrinsic components of the training process. Furthermore, the model is most effective in contexts without budgetary constraints or time pressure to meet deadlines (Holton & Naquin, 2005; Passmore & Velez, 2012). From the authors' perspective, striking a balance between the model's thorough evaluation approach and the practical constraints of real-world organisational settings is essential for its successful implementation.

2.3.6 Bushnell's input, process, and output model

In contrast to Kirkpatrick's model, which is typically applied immediately after a training event, Bushnell (1990) introduces a training evaluation model that can be employed before and after training, encompassing formative and summative assessment aspects (Bomberger, 2003). Bushnell's model, known as the Input, Process, Output (IPO) model, adopts a systemic approach encompassing the entire training process. Moreover, while Kirkpatrick's model needs to be revised to measure long-term financial outcomes, Bushnell's framework extends its assessment horizon to include extended periods, such as assessing factors like profitability, competitive enhancement, and business viability. This model articulates the value of training in financial terms, encompassing metrics like profitability, customer satisfaction, and productivity (Chang, 2010). It aids in determining whether a training program's intended objectives were achieved, the necessity for further program enhancements, and the acquisition of requisite knowledge and skills by trainees (Bushnell, 1990; Galvin, 1983; Phillips, 2000).

The Input, Process, Output model amalgamates elements from Kirkpatrick's first three levels - reaction, learning, and behaviour - with those from Brinkerhoff's six-stage model (1987), including the assessment of needs and objectives, program design, program operation, learning evaluation, usage and retention evaluation, and payoff evaluation (Chang, 2010). Bushnell contends that evaluation must transpire across every phase of the training system, encompassing components of input, process, and output. Each of these components involves quantifiable factors. The input phase evaluates variables such as learner qualifications, program design, trainer qualifications and competency, material quality, facilities, and equipment, all of which contribute to the efficacy of the training intervention (Passmore & Vele, 2012). The process phase entails scrutinising the developmental aspects of the plan, encompassing design, development, and program delivery (Jain, 2014). The output stage appraises trainees' reactions, knowledge acquisition, newly acquired skills, and enhanced job performance (Bushnell, 1990). Notably, this output stage mirrors Kirkpatrick's initial three levels: reaction, learning, and behaviour (Jain, 2014). The term "output" denotes short-term outcomes or the immediate impact of training (Bushnell, 1990; Jain, 2014). The outcome stage encompasses the longer-term effects of training, aligning with organisational enhancements such as profitability, customer satisfaction, and productivity (Bushnell, 1990; Jain, 2014).

However, this model exhibits certain limitations. It must provide adequate insights into program functioning and fully consider specific influencing factors contributing to these outcomes (Passmore & Velez, 2012). Furthermore, empirical observations based on proposed criticisms have not substantiated significant impacts or effects derived from this model (Robertson, 2004). Moreover, this model remains theoretical, with limited practical applications (Topon, 2012).

While each of the above models offers unique advantages, they also have specific limitations that require further development and refinement to enhance their applicability and effectiveness in evaluating training programs. Therefore, the current paper proposes using a sequential explanatory mixed methods design to improve the evaluation of the transfer of training for teachers.

3 METHOD

A comprehensive literature search was executed using the scoping review methodology to ascertain pertinent scholarly investigations concerning the transfer of training. This systematic inquiry encompassed peer-reviewed published articles and book chapters and spanned notable databases, such as Sage, NORA, ERIC, and Wiley Online Library. The primary objective of this endeavour was to comprehensively explore the existing body of literature pertinent to training transfer in a general context.

Adopting a scoping review is justified when an expansive corpus of evidence exists concerning an emerging subject of inquiry (Munn et al., 2018). This methodology facilitates the comprehensive delineation and mapping of foundational concepts and evidentiary underpinnings within a particular phenomenon (Whittemore et al., 2014) in the present case, the intricate domain of training transfer. The insights garnered through this preparatory stage can inform a systematic

literature review (Peterson et al., 2017) or catalyse the synthesis of forthcoming research endeavours (Arksey & O'Malley, 2005; Munn et al., 2018).

In contradistinction to the meticulous rigour inherent in a systematic literature review, the scoping review method affords latitude in encompassing diverse literature associated with the phenomenon in focus. These sources encompass an array of quantitative and qualitative journal articles, conference papers, and digital resources (Arksey & O'Malley, 2005). The process of conducting a scoping review involved the determination of search criteria for the literature, criteria for inclusion and exclusion, and collating and summarising the findings (Arksey & O'Malley, 2005). The findings could be presented in any way that could help the readers understand the phenomenon (Fung & Gunasekara, 2021).

3.1 Search Strategy and Search Criteria

Search Strategies focused on electronic searches of online databases such as NORA, ERIC, Wiley Online Library and Sage from 2021 to 2023 to determine the number of peer-reviewed articles related to the subject area. These databases include relevant journals in the field, such as the International Journal of Learning, Teaching and Educational Research, the International Journal of Educational Technology, the Journal of Research in Science Teaching, the Journal of Applied Social Psychology, the International Journal of Training and Development, the Journal of European Industrial Training, the International Journal of Social Research Methodology, and the Journal of mixed methods research.

The initial search of each database used general and inclusive search strings comprising the keywords "evaluation models" and "training transfer" to ensure the most considerable number of articles returned in this stage of the search strategy. The following table summarises the search results. It displays each database with the number of hits.

Table 1. Summary of search results.

NO	Databases	Number of Hits
1	Sage	702
2	NORA	648
3	ERIC	493
4	Wiley Online Library	315
	Total	2,158

3.2 Criteria for Inclusion and Exclusion

For those articles selected in the first search category, only articles in the English language and published in peer-reviewed academic journals were considered. All conceptual and empirical articles, both quantitative and qualitative, were included.

The search specified that the terms ‘‘training’’ and ‘‘transfer’’ appeared in the titles, abstracts, or keywords. This search generated more than 600 published book chapters or articles published or available as advanced online versions before the middle of February 2022. The dates searched were from 2000 to 2022. In the first round, all the articles or book chapters that included the exact term 'training transfer' in the title, abstract, keywords, or a measure of training transfer were retained, resulting in 200 published pieces. Twenty pieces were excluded because they were irrelevant to the main subject or the publication year before 2010. Another five were excluded because the author could not access the full text. This left 175 articles or chapters that were included in the review. Figure 1 shows a scoping review of PRISMA flow.

3.3 Collating, Summarizing and Reporting the Findings

At this stage, the selected literature was organised based on the themes (Arksey & O’Malley, 2005). The selected studies have been conducted in a wide range of Asian and European countries, as well as America and other countries such as South Africa, Australia, and Brazil. Studies were quantitative, qualitative, and mixed methods in nature. A summary was written on each piece of literature, and these summaries were then combined and reported in the findings section below.

4 FINDINGS

4.1 Training Evaluation

Training evaluation encompasses the delineation and measurement of specific outcomes to ascertain the organisational and trainee-level advantages engendered by training initiatives (Huang & Shen, 2021). Chirayu (2012) expounds upon the concept of training evaluation, characterising it as a systematic procedure involving the thorough gathering of descriptive and subjective information that is pivotal for making well-informed decisions about training, spanning facets such as selection, integration, value assessment, and customisation within the ambit of training initiatives. Conventionally, training evaluation is positioned as the concluding phase within the training continuum (Antoniou & Kyriakides, 2013). Nevertheless, Avalos (2011) contends that fundamentally reengineering the entire training paradigm is imperative. While conventional perspectives posit training as a reactionary response to performance impediments, a more optimal configuration entails its positioning as a pivotal intermediary course of action. Post-training, the concerted endeavours of managers, trainers, and trainees should coalesce towards the purposeful application of acquired training insights alongside diligent on-the-job management coaching (Alfonso & Ramirez, 2021).

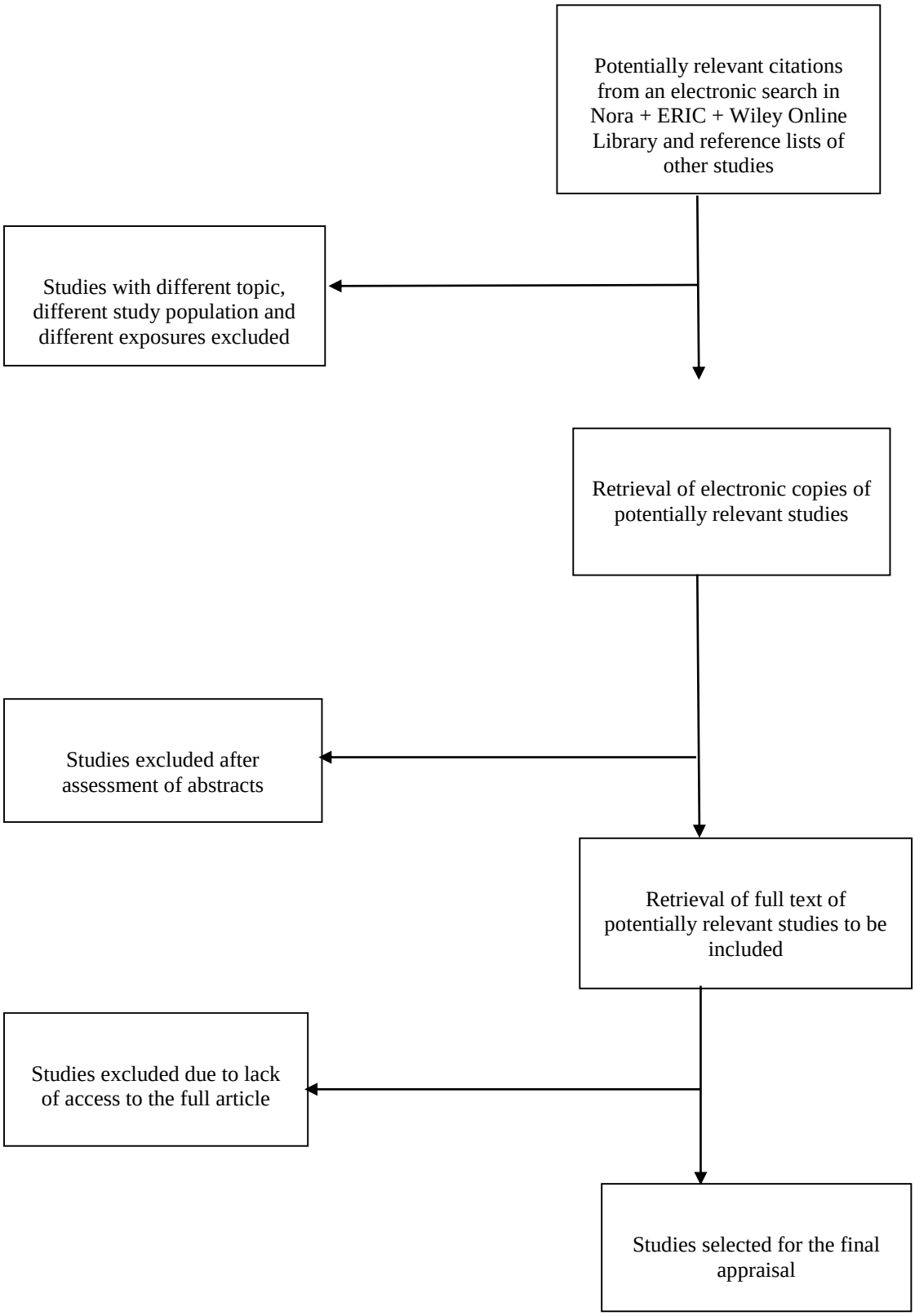

Figure 1. Scoping review PRISMA flow.

Aligned with this discourse, Devins and Smith (2013) underscore that the evaluation of training ought to manifest as an active and continuous endeavour spanning the entirety of the training trajectory. Evaluation should be intrinsically interwoven into the training program, rendering it an ongoing pursuit. A consortium of researchers accentuates the importance of systematically scrutinising training initiatives and delving into the intricate domain of training transfer while acknowledging the arduousness inherent in this stage (Wick et al., 2010; Kirwan, 2009; Al-Balushi, 2015). Furthermore, the tenor of the evaluation experience is poised for enhancement when it is woven into all junctures of the training journey, including the phase of training transfer (Al-Balushi, 2015).

4.2 Importance of Training Evaluation

Training evaluation is a critical facet of organisational learning and development, playing an indispensable role in assessing the effectiveness and efficiency of training interventions (Sales et al., 2021). This process involves systematically analysing training programs, their outcomes, and the extent to which they align with the established learning objectives. The significance of training evaluation stems from its capacity to inform evidence-based decision-making, enhance training design, optimise resource allocation, and contribute to achieving organisational goals (Devi & Shaik, 2012).

Ryan and Deci (2020) emphasise that the evaluation process allows trainers and organisations to gauge the extent to which participants can transfer learned knowledge and skills into real-world contexts, thereby establishing a link between the training room and workplace performance. This alignment is crucial to ascertain that the training intervention contributes meaningfully to organisational productivity and efficiency. Additionally, training evaluation assists in identifying areas that require improvement or modification in training design and delivery, fostering a continuous improvement cycle (Devi & Shaik, 2012).

Furthermore, training evaluation can contribute to efficient resource allocation by identifying programs that yield higher returns on investment (ROI). Organisations are better equipped to allocate resources judiciously when they possess empirical evidence of training's impact on employee performance and organisational outcomes (Phillips, 2016). This empirical foundation enables organisations to prioritise and tailor their training efforts, focusing on areas aligned with strategic objectives.

Therefore, training evaluation serves as an indispensable tool in the realm of organisational learning and development. By leveraging established evaluation models and principles, organisations can systematically assess training programs, measure their effectiveness, and align them with overarching organisational goals. This evidence-based approach enhances decision-making and ensures optimal utilisation of resources, fostering a culture of continuous improvement and innovation within the organisation ((Devi & Shaik, 2012).

4.2 Training Evaluation Models

A diverse array of training evaluation models has been developed over the preceding forty years (Passmore & Velez, 2012), with organisations adopting distinct models to gauge training efficacy (Topno, 2012). These training evaluation models are conceptualised to delineate the dimensions or factors that necessitate consideration when assessing training effectiveness (Tzeng et al., 2007). Broadly, these models can be categorised into two primary domains: goal-based frameworks, exemplified by Kirkpatrick's four-level model and system-based paradigms, such as the context, input, process, and product model (Phillips, 1991).

These two classification paradigms exhibit distinctive attributes. The goal-based approach facilitates delineating evaluative objectives from technical parameters to nuanced political considerations (Kennedy et al., 2014). It fosters the formulation of well-defined goals and strives to ascertain their attainment. Conversely, the system-based perspective expounds upon the sequential prerequisites to accomplish these goals and proffers strategies for utilising findings to refine training endeavours (Kennedy et al., 2014). This approach accentuates the effectiveness and efficiency of the intervention (Zinovieff & Rotem, 2008). The macro-level perspective concentrates on a specific training event, dissecting and elucidating its activities without explicitly encompassing the contextual elements influencing these activities. On the other hand, the micro-based framework delves into the internal and external organisational factors exerting influence on training activities (Madaus & Kellaghan, 2002).

Nevertheless, each of these evaluation models has its strengths and weaknesses, and their suitability depends on the specific context and goals of the training program. The following table summarises the different training evaluation models, focusing on strengths and areas that need further development in each model.

Table 2. Strengths and limitations of training evaluation models.

Model	Strengths	Limitations
Kirkpatrick's Four-Level Model	- Hierarchical structure guides evaluation process.	- Presumes causal relationships between levels.
	- Focuses on participant reactions, learning outcomes, behaviour change, and organisational results.	- Incompleteness in accounting for individual and contextual factors.
	- Emphasizes impact on organisational goals.	- Limited guidance on addressing contextual factors.
	- Widely adopted and recognized in training industry.	- Application may vary based on context and goals.

Kaufman and Keller's Five Levels of Evaluation Model	- Incorporates societal outcomes and value assessment.	- Require extensive data collection for ROI.
	- Expands Level 1 to include enabling and reaction.	- Theoretical foundation may limit practical use.
	- Addresses internal and external consequences.	- Not suitable for all training scenarios.
Phillips Return on Investment Model	- Focuses on economic value and ROI.	- Challenging to accurately compute ROI.
	- Provides evidence of training's financial benefits.	- Subjectivity in separating net benefits.
Warr et al's CIRO Model	- Incorporates contextual and input assessment.	- Doesn't require outcome assessment.
	- Sequential stages for holistic evaluation.	- Requires collaboration between stakeholders.
Brinkerhoff's Six-Stage Model	- Integrates formative and summative assessment.	- Time-consuming due to pre- and post-assessment.
	- Addresses training needs and design.	- Requires close collaboration between stakeholders.
Bushnell's Input, Process, Output Model	- Considers input, process, output, and outcomes.	- Doesn't fully account for program functioning.
	- Quantifiable aspects for comprehensive evaluation.	- Lacks evidence of significant impact in practice.

Combining aspects from different models or adapting them to fit the unique needs of a particular training initiative might be necessary for a comprehensive and practical evaluation.

5 WHAT NEEDS TO BE DONE IN FUTURE RESEARCH

Future research in training transfer evaluation should strive to create comprehensive and adaptable frameworks by synergising the strengths of existing models while addressing their limitations. The development of hybrid models that account for individual differences and contextual factors could offer nuanced insights into the dynamics of training transfer (Robertson, 2004). Longitudinal

assessment methods should be explored to capture the sustainability of transferred skills over time, and the refinement of practical implementation guidelines can enhance the feasibility of complex evaluation models (Brewer, 2007). Robust quantification of return on investment and budgetary impact necessitates innovative approaches like cost-benefit analyses while incorporating data-driven tools such as predictive analytics and machine learning, which can modernise evaluation processes and yield more accurate insights (Brewer, 2007).

To enhance context-specific applicability, action research and case study methodologies can provide in-depth insights into how different models perform in diverse industries and organisational contexts (Robertson, 2004). An efficient approach is the sequential explanatory mixed methods design, which combines quantitative and qualitative methods in two distinct phases. This design begins with quantitative data collection and analysis to identify trends and patterns, followed by qualitative data collection to provide a deeper understanding of these findings through participant perspectives and contextual nuances (Creswell & Clark, 2011). Mixed-methods designs can offer a more comprehensive understanding of the training transfer process, considering measurable outcomes and qualitative experiences (Creswell & Clark, 2011).

Participatory evaluation involving various stakeholders can provide holistic perspectives and align training evaluations with broader organisational objectives. Longitudinal studies and experimental designs can help establish causal relationships and uncover mediating factors influencing training transfer outcomes (Tashakkori & Creswell, 2007). Comparative studies directly comparing the effectiveness of different evaluation models in specific contexts can offer empirical evidence to guide practitioners in selecting the most suitable approach for their training programs. Surveys, interviews, and focus groups can also capture trainees' perceptions and attitudes, shedding light on factors influencing training transfer success (Tashakkori & Creswell, 2007). Collaborative evaluation platforms and online surveys can facilitate efficient data collection and analysis, making evaluation processes more streamlined and accessible. By embracing these diversified measurement tools and research methods, training transfer evaluation can evolve, providing practitioners with practical tools to assess training impact, enhance employee development, and foster overall organisational growth (Creswell & Clark, 2011).

This paper proposes a sequential explanatory mixed methods design to comprehensively address the gap with the previous models. In this approach, quantitative data is collected and analysed first, followed by the subsequent collection of qualitative data to enhance understanding of the research problem (Creswell, 2002; Harris, 2011). The rationale for proposing a sequential explanatory mixed methods design, where quantitative data is collected first, followed by qualitative data, needs to be justified. This design allows for a more comprehensive exploration of the research problem by combining numerical data and rich qualitative insights (Harris, 2011).

The rationale for employing a sequential explanatory mixed methods design lies in the understanding that specific research topics or events may be complex to capture using quantitative or qualitative methods alone comprehensively. By using a sequential explanatory mixed methods design, the limitations of each approach can be overcome, and the analysis can be enriched, resulting in a more nuanced and complete understanding of the research topic (Shannon-Baker, 2015; Brown, 2014; Johnson & Onwuegbuzie, 2004). Quantitative methods provide numerical

data that offer generalizability and statistical analysis, allowing for identifying patterns, correlations, and statistical significance. However, they may need more contextual depth and insights into the lived experiences of individuals involved in the research. On the other hand, qualitative methods offer in-depth, rich insights into the experiences, perceptions, and meanings associated with the research topic. They allow for a deeper exploration of individual perspectives and contextual factors. However, qualitative findings may lack generalizability and statistical rigour (Brown, 2014).

Employing a sequential explanatory mixed methods design can address each method's limitations (Shannon-Baker, 2015). The quantitative phase provides a broad understanding of the research topic, while the qualitative phase delves deeper into individuals' experiences and perceptions, offering a more comprehensive picture. Integrating quantitative and qualitative data allows for triangulation, validation, and more robust analysis, enhancing the overall quality of the study (Shannon-Baker, 2015; Brown, 2014; Johnson & Onwuegbuzie, 2004).

While some methodological purists argue against the use of mixed-method research, advocating for researchers to work exclusively within either a qualitative or quantitative paradigm (Bradbury et al., 2017), the sequential explanatory mixed-method approach offers a valuable opportunity to gain a more comprehensive understanding of the training transfer process (Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 2010; Morse & Niehaus, 2009). The purpose of proposing a sequential explanatory mixed-method design is to provide a more nuanced and complex understanding of the research topic, which may not be achieved through a single method alone. This approach embraces the idea that different methods can complement and enhance each other, allowing for a more comprehensive exploration of the research question (Flick, 2011). It is characterised by its pluralistic and inclusive nature, where the selection of methods is guided by what is needed to effectively address the research question (Creswell & Plano Clark, 2011). Top of Form.

6 CONCLUSION

Investigating why employees, as teachers choose to use skills and knowledge acquired during a training program, is crucial to understand the training transfer process better and identify why training transfer within an organisation succeeds or fails. The training evaluation process should be active and ongoing throughout the entire training cycle, including the training transfer process. The exploration of various training transfer evaluation models sheds light on the complexities and nuances inherent in assessing the effectiveness of training programs. Each model presents distinct strengths and limitations, emphasising the need for a holistic and adaptable approach to evaluating training transfer. The future of training transfer evaluation research lies in developing hybrid models that amalgamate the strengths of existing frameworks while mitigating their shortcomings. Such hybrid models should acknowledge individual and contextual factors, account for short- and long-term outcomes, and facilitate practical implementation. To elaborate more, a promising avenue for future research involves the development of hybrid models that draw from the strengths of various frameworks while actively addressing their shortcomings. The hybrid model represents an integrative approach, transcending the constraints of individual models to offer a more comprehensive and nuanced understanding of the training transfer process. In crafting such hybrid

models, it becomes imperative to incorporate a spectrum of individual and contextual factors that influence training transfer outcomes. These factors can encompass the diverse backgrounds, motivations, and learning styles of employees as teachers and the unique organisational contexts in which training programs unfold. By acknowledging and integrating these elements, a hybrid model can offer a more tailored and realistic assessment of the training transfer process.

Furthermore, a robust hybrid model should extend its purview beyond immediate outcomes, considering both short- and long-term impacts. This temporal dimension is crucial in capturing the sustainability and enduring effects of training on employee performance and organisational outcomes. Evaluating training transfer as an ongoing process rather than a one-time event ensures a more accurate representation of its effectiveness over time. Practical implementation stands as a cornerstone in the efficacy of any evaluation model. A well-designed hybrid model should be theoretically sound and pragmatic in its application within organisational settings. This entails a user-friendly design, clear guidelines for implementation, and compatibility with diverse organisational structures and cultures.

Moreover, incorporating innovative measurement tools and research methods can significantly enhance the accuracy and depth of training transfer evaluations. The introduction of robust quantification methods like return on investment (ROI) analysis can bridge the gap between training outcomes and financial impacts, providing a comprehensive view of training's value. Integrating data-driven tools like predictive analytics and machine learning can modernise evaluation processes and ensure data-driven decision-making.

The sequential explanatory mixed methods design is a prime example of a research approach combining quantitative and qualitative methods, allowing for a more comprehensive understanding of training transfer. The sequential explanatory mixed methods design is supported by its inherent ability to offer a comprehensive view of research phenomena by merging quantitative and qualitative data. Numerous studies across diverse disciplines have successfully employed this design, highlighting its effectiveness in uncovering nuanced insights that might be overlooked by employing only one method. By capitalising on the strengths of both data types, researchers can uncover trends, patterns, and underlying participant perspectives, thereby enriching the evaluation process. Prominent research methodology guides and textbooks emphasise the value of the explanatory mixed methods design. Authors such as Creswell and Plano Clark (2018) and Creswell and Creswell (2017) discuss its merits and provide step-by-step guidance. The sequential explanatory mixed methods design starts with the quantitative data through collecting numerical data. This could involve surveys, experiments, or other quantitative methods followed by analysing the quantitative data using appropriate statistical methods to identify patterns, trends, and relationships. After that, the qualitative stage starts by collecting data, ensuring it complements and enriches the quantitative findings, followed by analysing the qualitative data using thematic analysis, coding, or other qualitative methods to uncover underlying participant perspectives. Saying this, the process of comparing the quantitative and qualitative findings will take place to identify converging or diverging patterns and develop an integrated interpretation that provides a more holistic understanding of the research question.

By revealing gaps in current knowledge and identifying previously unknown areas for improvement in training transfer, this paper contributes novel insights to the field of training transfer. In embracing these recommendations, the field of training transfer evaluation can move towards a more comprehensive, adaptable, and data-informed paradigm. This evolution will equip practitioners with the tools to optimise training programs, enhance employee performance, and foster sustainable organisational growth.

ACKNOWLEDGEMENTS

This research received no specific grant from public, commercial, or not-for-profit funding agencies.

REFERENCES

Alfonso, M., & Ramirez, R. (2021). The impact of professional learning communities on teacher professional development and student achievement: A meta-analysis. *Educational Research Review, 36,* 100366.

Ajzen, I., & Fishbein, M. (1977). Attitude-behaviour relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84(*5), 888–918.

AL-Aufi, M. (2014). Implications of the Improvement of Teaching Quality for Professional Development (PD) of Academics at the Colleges of Applied Sciences (CASs) in the Sultanate of Oman (PhD Thesis). University of Waikato, New Zealand.

AL-Balushi, K. (2012). Why are teachers reluctant to provide professional development? In N. McBeath & N. AlKalbani (Eds.), Quality in ELT: Raising Pedagogical Standards (pp. 87–97, ELT Conference proceedings). Muscat: SQU.

AL-Balushi, K. (2015). Internet-based surveys: How can they be utilised to research in education? In Proceedings of INTED2015 Conference (pp. 0662–0670). Madrid, Spain: INTED.

AL-Ghatrifi, Y. (2016). The Professional Development of Teachers in Higher Education in Oman: A Case Study of English Teachers in the Colleges of Applied Sciences (PhD thesis). University of Reading, United Kingdom.

AL-Hakamani, F. (2011). Investigating the Collaborative Professional Practices of Omani EFL Teachers and Their Relation to the Quality of Their Reflection (MA Dissertation). Sultan Qaboos University, Muscat.

AL-Jardani, K. (2015). The Need for Bridging the Gap between General and Tertiary Education in Oman. *English Linguistics Research, 4*(2), 45-48.

Al Jabri, M., Silvennoinen, H., & Griffiths, D. (2018). Teachers' professional development in Oman: Challenges, efforts and solutions. *International Journal of Learning, Teaching, and Educational Research, 17*(5), 82–103. https://doi.org/10.26803/ijlter.17.5.6.

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology, 42,* 331-342.

Antoniou, P., & Kyriakides, L. (2013). A Dynamic Integrated Approach to teacher professional development: Impact and sustainability of the effects on improving teacher behaviour and student outcomes. *Teaching and Teacher Education, pp. 29*, 1–12.

Avalos, B. (2011). Teacher professional development in Teaching and Teacher Education over ten years. *Teaching and Teacher Education, 27,* 10–20.

Antoniou, P., & Kyriakides, L. (2013). A Dynamic Integrated Approach to teacher professional development: Impact and sustainability of the effects on improving teacher behaviour and student outcomes. *Teaching and Teacher Education, 29*, 1–12.

Al-Omairi, A. (2021). The Reasons Behind Not Passing the Requirements of the New Teachers Program of the Specialised Institute for Professional Training of Teachers. ResearchGate/12816/0055865.

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. International *Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616.

Arthur, W., Jr., Bennet, W., Edens, P.S., & Bell, S.T. (2003). Effectiveness of training in organisations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology, 88*(2), 234-245.

Baldwin, T. T., & Ford, J. K. (2017). Transfer of training: The known and the unknown. *The Annual Review of Organisational Psychology and Organisational Behaviour, 5,* 201-225.

Bates, R. A. (2001). Public sector training participation: An empirical investigation. *International Journal of Training and Development, 5*(2), 136–150.

Bates, R. (2004). A critical analysis of evaluation practice: The Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning, 27*, 341–347.

Barnett, B.G. (2005). Transferring learning from the classroom to the workplace: Challenges and implications for educational leadership preparation. *Educational Considerations, 32*(2), 6-16.

Bendall, T. C., & Sanders, K. (2017). Do Opportunities for Formal Learning Stimulate Follow-up in Formal Learning? A three-wave study. *Human Resource Management, 56,* 803–820.

Brewer, E. W. (2007). Toward a holistic evaluation framework: A synthesis of alternative approaches to evaluation. *American Journal of Evaluation, 28*(1), 64–78.

Brewer, T. (2007). Use of Phillips‘s five-level training evaluation and return-on-investment framework in the U.S. non-profit sector, doctoral dissertation—University of North Texas, Denton, TX.

Brinkerhoff, R. O. (1987). Achieving results from training. San Francisco, CA: Jossey-Bass.

Brinkerhoff, R. O. (1988). An integrated evaluation model for HRD. Training and Development Journal, 66-68.

Brinkerhoff, R. O. (2003). The success case method. San Francisco: Berrett-Koehler.

Burke, L., & Hutchins, H.M. (2008). Training transfer: An integrative literature review. *Human Resource Development Review, 6*(3), 263–296.

Bushnell, D. S. (1990). Input, process, output: A model for evaluating training. *Training and Development Journal, 44*(3), 41–43.

Cannon-Bowers, J. A., Salas, E., Tannenbaum, S. I., & Mathieu, J. E. (1995). Toward theoretically based principles of training effectiveness: a model and initial empirical investigation. *Military Psychology, 7*, 141-164.

Chen, J., & Hou, J. (2021). The influence of sustainability on training transfer: Evidence from China's manufacturing industry. *Journal of Business Research, 131*, 128-139. http://dx.doi.org/10.1016/j.jbusres.2021.02.044.

Chiaburu, D. S., Van Dam, K., Hutchins, H. M., & (2012). Social Support in the Workplace and Training Transfer: A Longitudinal Analysis. *International Journal of Selection and Assessment, 18*(2), 186–201.

Clemenz, C.E. (2001). Measuring Perceived Quality of Training in the Hospitality Industry. Doctoral thesis, The Faculty of the Virginia Polytechnic Institute and State University, USA.

Creswell, J. W. (2008). Educational research: Planning, conducting, and evaluating Qualitative and Quantitative research. New Jersey: Merrill Prentice Hall.

Creswell JW, Plano Clark VL (2011). Designing and conducting mixed methods research. 2nd ed. ThousandOaks, CA: Sage.

Creswell, J. W., & Plano Clark, V. L. (2017). Designing and conducting mixed methods research. Sage Publications.

Devi, R. V., & Shaik, N. (2012). Evaluating training & development effectiveness - A measurement model. *Asian Journal of Management Research, 2*(1), 722-735.

Devins, D., & Smith, J. (2013). Evaluation of human resource development. Gold, J., Holden, R., Stewart, J., Iles, P., Beardwell, J (Eds). Human resource development: Theory and practice, chapter 7, pp. 175–204, Basingstoke: Palgrave Macmillan.

Flick, U. (2011). Introducing Research Methodology: A Beginner's Guide to Doing a Research Project. London: Sage Publications.

Ford, J. K., Quinones, M. A., Sego, D. J., & Sorra, J. (1992). Factors affecting the opportunity to perform trained tasks on the job. *Personnel Psychology, 45*(3), 511–27.

Ford, J. K., & Kraiger, K. (1995). The application of cognitive constructs to the instructional systems model of training: implications for needs assessment design and transfer. In Cooper, C.L. and Robertson, I.T. (Eds), International review of industrial and organisational psychology, 1-48, Chichester: Wiley.

Foxon, M. (1989). Evaluation of training and development programs: A review of the literature. *Australian Journal of Educational Technology, 5*(2), 89–104.

Grossman, R., & Salas, E. (2011). The transfer of training: what matters. *International Journal of Training and Development, 15*(2), 103-120.

Hamblin, A. C. (1974). Evaluation and control of training. Maidenhead: McGraw-Hill.

Harrits, G.S. (2011). More than method?: A discussion of paradigm differences within mixed methods research. *Journal of Mixed Methods Research, 5*, 150-166.

Holton, E. F. III, Bates, R. A., Seyler, D. L., & Carvalho, M. B. (1997). Toward construct validation of a transfer climate instrument. *Human Resource Development Quarterly, 8*(2), 95–113.

Holton, E. F. III, Bates, R. A., & Ruona, W. E. (2000). Development of a generalised Learning Transfer System Inventory. *Human Resource Development Quarterly, 11*(4), 333–60.

Holton III, F. E. (1996). The Flawed Four-Level Evaluation Model. *Human Resource Development Quarterly, 7*(1), 5–21.

Huang, C., & Shen, H. (2021). The impact of teacher collaboration on student achievement: A meta-analysis. *Educational Research Review, 32,* 100373.

Hung, T. (2010). An empirical study of the training evaluation decision-making model to measure training outcome. *Social Behaviour and Personality: An International Journal, 38*, 87–102.

Jamjoom, M.M., & Al-Mudimigh, A.S. (2011). Training Evaluation: Towards an Effective ES Training. *International Journal of Computer Science and Network Security, 11*(1), 148–152.

Johnson, R., & Onwuegbuzie, A. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher, 33*(7), 14-26.

Joffe, H. (2012). 'Thematic analysis.' Qualitative Research Methods in Mental Health and Psychotherapy (eds D. Harper and A. R. Thompson). http://dx.doi.org/10.1002/9781119973249.ch

Kaufman, R., Keller, J., & Watkins, R. (1996). What works and what doesn‘t: Evaluation beyond Kirkpatrick. *Performance + Instruction, 35*(2), 8–12.

Kennedy, P. E., Chyung, S. Y., Winieck, D. J., & Brinkerhoff, R. O. (2014). Training professionals‘ usage and understanding of Kirkpatrick‘s level 3 and level 4 evaluations. *International Journal of Training and Development, 18*(1), 1-21.

Khalid, M., Rehman, C., & Ashraf, M. (2012). Exploring the link between Kirkpatrick (KP) and context, input, process and product (CIPP) training evaluation models and its effect on training evaluation in public organisations of Pakistan. *African Journal of Business Management, 6*(1), 274–279.

Kim, Y. J., Lee, H. J., & Kwon, H. J. (2020). Examining the role of teacher motivation in student learning outcomes: A cross-sectional study. *Asia Pacific Education Review, 21*(3), 401-413.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). Evaluating training programs: The four levels (3rd ed.). San Francisco: Berrett-Koehler Publication.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of Cognitive, Skill-Based, and Affective Theories of Learning Outcomes to New Methods of Training Evaluation. *Journal of Applied Psychology, 78*(2), 311-328.

Lockwood, S. L. (2001). Enhancing employee development: Development and testing of a new employee orientation protocol. Doctoral thesis, California School of Professional Psychology, San Diego.

McKenna, E. F., & Beech, N. (2014). Human resource management: A concise analysis. New York: Pearson Education.

Milne D.L. (2007). Evaluation of staff development: The essential ‗SCOPPE. *Journal of Mental Health, 16*, 389–400.

Morse, J., & Niehaus, L. (2009). Mixed methods design: Principles and procedures. Walnut Creek, CA: Left Coast Press.

Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology, 18*(1), 1–7. https://doi.org/10.1186/s12874-018-0611-x.

Nickols, F. W. (2005). Why a stakeholder approach to evaluating training. *Advances in Developing Human Resources, 7*(1), 121-134.

Noe, R. (2016). Employee Training and Development. Boston: Irwin/McGraw-Hill.

Passmore, J., & Velez, G. A. (2012). International coaching psychology review: A bibliography of research using coaching psychology themes by academic institutions and professional bodies. *International Coaching Psychology Review, 7*(1), 37-50.

Peterson, J., Pearce, P. F., Ferguson, L. A., & Langford, C. A. (2017). Understanding scoping reviews: Definition, purpose, and process. *Journal of the American Association of Nurse Practitioners, 29*(1), 12–16. https://doi.org/10.1002/2327-6924.12380.

Phillips, J. J. (1996). How much is the training worth? *Training and Development, 50*(4), 20–24.

Phillips, J. J. (2005). Investing in your company's human capital: Strategies to avoid spending too little or too much. New York: Amacom.

Preskill, H., and Russ-Eft, D. (2005). Building evaluation capacity: 72 activities for teaching and training. Thousand Oaks: Sage.

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, 101860.

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.

Robertson, P. J. (2004). Beyond Return on Investment (ROI) and Kirkpatrick: A guide to evaluating training. *Journal of European Industrial Training, 28*(2/3/4), 23–31.

Roca-Puig, V., & Llorens-Montes, F. J. (2021). How corporate sustainability affects the transfer of training: The mediating role of organisational learning capability. *Journal of Business Research, 124,* 154-166. https://doi.org/10.1016/j.jbusres.2021.01.022.

Russ-Eft, D. F., Preskill, H., & Sleezer, C. (1997). Human resource development review: Research and implications. Thousand Oaks, CA: Sage Publications.

Saks, A. M. and 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), 118-12.

Salas, E. & Cannon-Bowers, J. A. (2001). Reflections on shared cognition. *Journal of Organisational Behaviour, 22*, 195–202.

Salas, E., Tannenbaum, S. I., Kraiger, K., Smith-Jentsch, K. A., & Cannon-Bowers, J. A. (2021). The science of training and development in organisations: What matters in practice for 2021 and beyond. *Psychological Science in the Public Interest, 22*(2), 49-113.

Shannon-Baker, P. (2015). Making Paradigms Meaningful in Mixed Methods Research. *Journal of Mixed Methods Research, 10*(4), 319–334. http://dx.doi.org/10.1177/1558689815575861.

Stewart G.L. & Brown K.J. (2011). Human Resource Management Linking Strategy to Practice. Danver: John Wiley & Sons.

Smith, J., & Johnson, A. (2023). Rethinking Kirkpatrick's model: A critique and proposal for a comprehensive evaluation framework. *Human Resource Development Review*, *28*(4), 457–473.

Stokking, K. (1996). Levels of Evaluation: Kirkpatrick, Kaufman and Keller, and Beyond. *Human Resource Development Quarterly, 7*(2), 179–183.

Stufflebeam, D. L. (1983). The CIPP model for program evaluation. In Madaus, G. F., Scriven, M., and Stufflebeam, D. L. (Eds.), Evaluation models, 117–141. Boston, MA: Kluwer-Nijhoff.

Tamkin, P., Yarnall, J., and Kerrin, M. (2002). Kirkpatrick and Beyond: A review of models of training evaluation. Brighton: Institute of Employment Studies, Report 392.

Tannenbaum, S. I. and Yukl, G. (1992). Training and development in work organisations. *Annual Review of Psychology, 43,* 399-441.

Tashakkori, A., & Creswell, J. W. (2007). The new era of mixed methods. *Journal of Mixed Methods Research, 1*(1), 3-7.

Teddlie, C., & Tashakkori, A. (2009). Foundations of mixed-methods research: Integrating quantitative and qualitative approaches in the social and behavioural sciences. Thousand Oaks, CA: Sage.

Tennant, Ch., Boonkrong, M., and Roberts, P.A.B. (2002). The design of a training programme measurement model. *Journal of European Industrial Training, 26*(5), 230-240.

Topno, H. (2012). Evaluation of training and development: An analysis of various models. IOSR *Journal of Business and Management, 5*(2), 16–22.

Tzeng, G.H., Chiang, C.H., and Li, C.W., (2007). Evaluating intertwined effects in e-learning programs: a novel hybrid MCDM model based on factor analysis and DEMATEL. *Expert Systems with Applications, 32,* 1028–1044.

Wang, G. G., & Wilcox, D. (2006). Training evaluation: Knowing more than is practised. *Advances in Developing Human Resources, 8*(4), 528–539.

Warr, P., Bird, M., & Rackham, N. (1970). Evaluation of Management Training. London: Gower Press.

Watkins, R., Leigh, D., Foshay, R., & Kaufman, R. (1998). Kirkpatrick Plus: Evaluation and Continuous Improvement with a Community Focus. *Educational Technology Research & Development*, *46*(4), 90–96.

Wick, C. (2003). Going beyond the finish line: Keep them learning long after the training ends. *Training and Development, 57*(12), 17–19.

Wick, C.W., Pollock, R.V., & Jefferson, A. (2010). The six disciplines of breakthrough learning: How to turn training and development into business results. San Francisco, CA: Pfeiffer.

Whittemore, R., Chao, A., Jang, M., Minges, K. E., & Park, C. (2014). Methods for knowledge synthesis: an overview. *Heart & Lung*, *43*(5), 453–461. https://doi.org/10.1016/j.hrtlng.2014.05.014

Yamkovenko, V. and Holton, E. F. (2010). Toward a theoretical model of dispositional influences on transfer of learning: A test of a structural model. *Human Resource Development Quarterly, 21, 381- 410.*

Zinovieff, M.A. and Rotem, A. (2008). Review and Analysis of Training Impact EvaluationMethods, and Proposed Measures to Support a United Nations System Fellowships Evaluation Framework., Geneva, Switzerland: WHO‘s Department of Human Resources for Health.