Frameworks for Evaluating Education Programmes and Work Related Learning: A Scoping Review

Abstract:

Purpose
It has become accepted practice to include an evaluation alongside learning programmes that take place at work, as a means of judging their effectiveness. There is a tendency to focus such evaluations on the relevance of the intervention and the amount of learning achieved by the individual. The aim of this review is to examine existing evaluation frameworks that have been employed to evaluate education interventions and in particular assess how these have been used and the outcomes of such activity.

Design/methodology/approach
A scoping review of the literature was undertaken to examine existing evaluation frameworks claiming to evaluate education interventions, and in particular to assess how these have been used and the outcomes of such activity.

Findings
Forty five articles were included in the review. The majority of papers concentrate on learner satisfaction and/or learning achieved. Rarely is a structured framework mentioned, or detail of the approach to analysis cited. Typically, evaluations lacked baseline data, control groups, longitudinal observations and contextual awareness.

Originality
This scoping review is novel in its assessment and critique of evaluation frameworks employed to evaluate work-related education programmes.

Practical implications
This review has implications for those involved in designing and evaluating work-related education programmes as it identifies areas where evaluations need to be strengthened and recommends how existing frameworks can be combined to improve how evaluations are conducted.

Keywords: evaluation, education programmes, work-related learning, scoping review

Introduction

Over the past two decades, and with the growth of learning interventions that take place, and/ or are closely related to the workplace, interest has arisen around the degree to which such interventions achieve their desired impact on performance. For example, we have seen this with the introduction of work-based leadership programmes, quality improvement (QI) education interventions to name but a few. For these and other workplace education interventions, it has become accepted practice to conduct an evaluation alongside a programme as a means of judging its effectiveness (Derwen, 2012). There is a tendency to focus such evaluations on the relevance of the intervention and the amount of learning achieved by the individual. Less importance is paid to the context within which the learning intervention is situated, and thus its impact on the workplace. But the organisational culture and systems related to the workplace are of great significance and will influence the success or otherwise of the intervention (Stufflebeam, 2005).

Evaluations are often introduced late in the life of a programme. Frequently with a narrow focus on the individual rather than those relating to the two-way influence brought on by the organisational setting within which they occur. This results in a strong focus on outcomes at the expense of process. At the same time there is often a lack of a "theory of use"; unawareness of a dichotomy between what is intended and believed to happen and what actually happens (see Argyris & Schon 1974). One of the key reasons for growing interest in theory-driven evaluation is increasing recognition of the inability of even the most sophisticated learner-focused evaluations to explain which factors are responsible for a programme's success or failure. Broadening the scope of an evaluation to include context and cultural surroundings becomes imperative.

What appears to be missed is engagement with the large body of work published over the last fifty years relating to how people learn. Building on a foundation provided by Knowles on andragogy (1980) and Mezirow (1997) on transformative learning with antecedent contributions from Dewey (1963) and Lewin (1973) on experiential learning and Kolb on reflective learning (1984). Attention has been drawn to the significance of individual responsibility related to adult learning and the way that this is embedded in action (see Revans, 1982). Equally important to this body of theories is the social and contextual element and the way that we learn from others (see for example Vygotsky (1978) on legitimate peripheral participation and Wenger on Communities of Practice (1998)).

More recent interest in work related learning, building on adult learning theories, has seen outstanding contributions from several scholars including: Eraut (2002) on professional competence; Billett (2005) on workplace pedagogy; Billett et al., (2012) on practice-based education; Unwin and Fuller (2005) relating work organisation to learning opportunities; and the huge and ongoing body of work by Engestrom (2016) in Finland on expansive learning.

Evaluations that focus on the participant experience alone tend to miss the significance of the context within which a learning intervention is situated. Providing a two dimensional rather

than a three-dimensional picture. A second area of weakness relates to terminology and its interpretation. Use of terms such as "effectiveness" and "impact" pre-supposes that we have common understanding of meanings which is often far from the case. As Horwich (2004) demonstrates, the way we use common terms is related directly to the context and the culture within which we operate. Misperceptions arise when beliefs are not articulated resulting in a range of starting and therefore end points. If the purpose of a learning intervention is not clear and is not shared at the outset then the way that it is judged will be variable. If one stakeholder believes that the purpose of a learning intervention is to have outcomes that relate to the individual whilst another stakeholder has organisational benefits in mind. Then judgements about effectiveness and impact will be different.

Many evaluations of learning interventions described in the literature look primarily at the learner perspective and are judged to be providing a positive learning experience for participants and thus the intervention is deemed to have been successful. Whilst positive learning outcomes are laudable, the organisational results are of at least equal importance but frequently less clear. A methodology is sought by which a holistic summary of strengths and weaknesses of an intervention and its impact can be judged. With these things in mind, a scoping study of the literature has been performed in order to identify a structured method of evaluating a programme of learning. The aim of this review is to examine existing evaluation frameworks that have been employed to evaluate education interventions and in particular assess how these have been used and the outcomes of such activity.

This paper explores those models that might be employed to support a theory driven evaluation. The majority of these models draw attention to the need to be clear about intentions related to terminology. For the purposes of this paper we define "effectiveness" to mean a learning intervention which fulfils its intended purpose and "impact" to mean a demonstrable difference to the manner in which work is performed that can be related directly to the content of the learning intervention.

For clarity we offer the following table of keywords and their use in this paper.

Insert table 1 about here

Methodology

The use of scoping reviews to examine the breadth of literature relating to a selected topic is becoming increasingly popular (Arksey and O'Malley, 2005). This method of analysis was selected over more traditional systematic reviews, as a scoping review permits cumulative identification and evaluation of literature, employing a number of approaches, to identify existing gaps in literature and thus to inform current debate and a future research study.

Although there is no universally accepted definition of a scoping review, or an exacting set of procedures (Peters et al., 2017). Arksey and O'Malley (2005) were the first to publish a five-stage framework detailing the purpose of this method. The scoping review for this study was used as a stand-alone initiative to draw conclusions and identify gaps in the existing literature

(Arksey and O'Malley, 2005). The methodology is similar to systematic activities completed in any review, including focus on a specific topical area; a well-defined research question; rationale regarding inclusion and exclusion criteria; and clearly defined procedures (Peterson et al., 2017).

The five stages of this scoping framework consist of: Identifying the research question, identifying relevant studies, study selection, charting the data and collating, summarising, and reporting the results (Arksey and O'Malley, 2005, p.8-9). Use of this method allowed credible and reflexive examination of a wide variety of literature on evaluation frameworks. Along with ensuring methodological rigour and transparency, enabling future replication and validation of the current findings.

As guided by the framework, our starting point was to ask the underlying research question: *"What is known from the existing literature about the effectiveness of a range of models/ frameworks that purport to evaluate the impact of learning programmes?"*

We believe that any evaluation should be comprehensive and multi-dimensional, capturing all elements of an intervention, should capture all relevant data, and should provide a clear strategy for analysis of findings, thus providing a comparable and credible evaluation pathway. Such a model would effectively measure process and outcomes / impact of an intervention designed to increase professional learning and organisational efficiency, depending on the stated objectives of the intervention. We recognised ambiguities found in the research question including what is meant by "effectiveness"; the definition of "impact" and what constitutes a "learning programme". We needed to be clear about what we mean by a model/framework and why this is significant. After spending some time clarifying these points (also see table 1) we approached the literature to establish the extent to which this comprehensive evaluation approach was being employed (second and third stages of the Arksey and O'Malley (2005) framework). Defining the search strategy also required careful consideration and deliberation among the research team. It was agreed the search needed to focus on work-based learning programmes rather than generic terms such as training and development which we felt were too broad for the scope of our review.

Our database search (see table 1) revealed 276 articles of potential relevance (see figure 1), judged by the search terms used. These were accessed and 73 duplicates were removed. Abstracts were read of the remaining 203 papers, leading to a selection of 95 appearing to match our inclusion criteria. This group was reviewed and skim read to assess their relevance relating to the criteria, and it was agreed that 45 articles matched the inclusion criteria sufficiently to make them worth closer scrutiny. Several of the articles focused on a review of the literature and did not describe a case study, but it was felt that they would provide useful information and so were included in the close reading stage. Citation tracking of reference lists led to a more detailed reading of a minority of the frameworks described. Exclusions were made because of lack of theoretical underpinning or a sketchy description of the approach to data collection and analysis.

Insert figure 1 about here

Insert table 2 about here

Findings - collating, summarising and reporting the results

The final two stages of Arksey and O'Malley's framework require the results of the study to be charted and then collated and reported. First we collate and summarise the papers included. Then follow this with a more detailed discussion of the papers included in this review.

Although it is not difficult to find articles that claim to include an evaluation strategy most of them provide, at best, a superficial picture of focus and conduct of the evaluation they describe. The majority concentrate on learner satisfaction and / or learning achieved. Rarely is a structured framework mentioned, or detail of the approach to analysis cited. As Trojan et al., (2009) point out there is generally a lack of baseline data, control groups, longitudinal observations, contextual awareness.

Of the papers closely surveyed for this report, 23 papers (50%) described a specific approach to data collection and analysis, 55% cited a framework for analysis (some a known example whilst others were created for the evaluation); 80% tested using a case study example of the evaluation methodology employed, and 50% critiqued their methodology. Although the majority of the evaluations described were located in the health sector, some were conducted in other locations, e.g. teaching, aviation, social work, which suggests that the limitations of evaluation technique are not unique to healthcare.

Broadly the papers fall into three categories; firstly there are a large number of papers that claim to have evaluated an intervention but describe simply the conduct of questionnaires and / or interviews and the drawing of conclusions from these (see for example Lechner et al 1999). Secondly there are a number of papers that claim to have used a framework but do not provide details of how and why they used it. (see for example Gallivan et al., 2008). Thirdly there are papers that claim to have used an approach but the explanation appears partial (see for example Thies et al., 2007). Of the known frameworks mentioned, Kirkpatrick is by far the most frequently cited; others include those of Donabedian (2005); Stufflebeam (2001), Pawson and Tilley (1997), Stern (2016), Stake (1967), Jarvis (1987). The Kirkpatrick model (2006), which consists of four levels – satisfaction, learning, behaviour and impact - is generally considered to have been developed as an evaluation method to critique learning interventions, but there are also several review papers that have used Kirkpatrick as an ordering / categorisation device. For the remainder of this paper we will focus on the papers that describe an evaluation approach that can be replicated.

For example Campbell et al., (2017) conducted a systematic review of the literature relating to Cancer Education for Nurses. They used the Kirkpatrick model as their ordering device

and found that two thirds of the papers (of 30) had no underpinning theory and several papers looked at level 1 only. There was evidence of awareness of underpinning theories, including adult learning theory, interpretative pedagogy, constructivism, active learning, reflection, whilst one cited situated learning, legitimate peripheral participation, and Communities of Practice. But evidence of transition from awareness of theory to influence on evaluation was scant. They also found that the quality of statistical analysis was generally weak and the choice of outcome measures frequently problematic and often opportunistic. In summary they pointed out that the majority of papers described their evaluation at level 1 (learner satisfaction) with little regard for context or outcomes. However they point out that

"it is acknowledged that trying to find a causal link between the education and the clinical outcome data is challenging because there is a sense that, in these studies, the research is trying to move this educational evidence base forward". Campbell et al., (2017, epub).

Johnston et al., (2018) undertook a review of the literature with regard to simulation and de-briefing in medical education using Kirkpatrick to examine how many papers cited differing levels. They found no papers giving results for levels 3 and 4. They point out that although levels 1 and 2 evaluations are not without value, the manner in which simulation and debriefing affects learning, behaviours, and ultimately patient outcomes is of great significance and not covered at these levels. Few studies have examined the true impact of simulation and debriefing as evidenced by Kirkpatrick's (2006) level 3 (changes in behaviour) and level 4 (results, or impact on patient outcomes). They urge researchers to support the continued maturation of the simulation pedagogy by aspiring to all levels of Kirkpatrick evaluation.

As an approach to reviewing, sorting and judging evaluation claims, Kirkpatrick has proved useful. But partial use of Kirkpatrick leaves evaluation questions unanswered. Trojan et al., (2009) attempt to overcome the narrow approaches generally found in the literature by proposing a broader evaluation approach looking at process / outcomes/ context / systems components in addition to the learning achieved, and by taking a mixed methods approach combining quantitative and qualitative data. They say that

"we would argue that adding process, context and systems evaluation components is essential for fully understanding how new practice models can be implemented and how they might affect patient outcomes. Trojan (2009 p.386)

Furthermore, in environments such as healthcare that are complex and constantly changing, it is difficult to conduct ''clean'', randomized, controlled interventions. There are typically many factors besides the study intervention that could affect outcomes and it is nearly impossible to attribute clear cause and effects." Trojan (2009 p.387)

Acknowledging that there are multiple factors that impact on the outcome of a learning intervention they suggest that

"the application of system theory in the evaluation of complex and dynamic settings such as healthcare is advantageous and should be explored further" Trojan (2009 p.387)

Freeth et al., (2010) also encourage a broad approach when assessing interprofessional education. Thistlethwaite et al., (2015) concur, suggesting that context + mechanism result in a holistic picture. Abma and Stake (2014) take a naturalistic view of each unique environment within which an intervention occurs. Therefore, although varying slightly in terminology and emphasis the consensus seems to be that a comprehensive evaluation should pay regard to outcomes, process, context and systems. Not only Trojan, but Donabedian and Stufflebeam also categorise around these terms.

The CIPP model associated with Stufflebeam (2001) resonates with the concept of context and mechanism evaluation, citing the need to address context, inputs, process and product evaluations. By this he means context evaluation to include examining and describing the organisational context of the intervention, conducting a needs and goals assessment, determining the objectives of the intervention. Input evaluation examines the inputs and resources and compares them with similar interventions. Process evaluation describes how the intervention was implemented and who was involved and what problems were experienced. Product evaluation includes determining and examining the general and specific outcomes of the intervention, measuring anticipated outcomes, trying to identify unanticipated outcomes, assessing the benefit of the intervention, conducting cost effectiveness assessment.

Donabedian's (2005) three components approach for evaluating the quality of care underpins measurement for improvement and is thus closely associated with the organisational context underpinning any learning intervention. The three components are structure, process and outcomes. Donabedian believes that structure measures have an effect on process measures, which in turn affect outcome measures. Together these form the basis of what is required for an effective suite of measures. The model has been critiqued widely, with particular reference to its linear nature, but it nevertheless has adherents amongst evaluators. Other authors suggest that cause and effect are more complex, particularly within the NHS where there is so much variability in contexts.

A framework that appears in only one paper of our survey but which warrants further examination is attributed to Stern (2016). This model, by highlighting the varying elements that make up the global picture of intervention creation, reminds us of context and its juxtaposition with problem identification, needs assessment, objectives, educational strategies, implementation and feedback. In this way we move away from a one dimensional evaluation of learner satisfaction and are drawn towards a view of the factors that accompany and impinge on learning.

Stake's (1967) framework also draws attention to the initial stages of any intervention. This model has two distinct aspects to it – a formative element, describing the process under scrutiny; and a summative element judging the effectiveness of the process. The Stake model

highlights the rationale that underpins any development, alongside a matrix marked Description and a comparable matrix marked Judgement. Each is then subdivided into three foci, examining the antecedents, transactions and outcomes of an intervention under scrutiny.

The model enables a reading of summary data and thus a holistic analysis to be performed. Analysis can be done by "reading" vertically or horizontally – ie either the intentions can be reviewed in terms of their antecedents, their transactions and their outcomes; followed by observations of what actually occurred and what the outcomes were. Alternatively the model can be read horizontally, so that the antecedents can be reviewed in terms of intentions, observations, standards and judgements. This latter enables a view of the congruence between each element, i.e. how closely intentions were matched by reality, and by accepted standards.

Jarvis (1987) highlights the social nature of the individual and their learning style (see also Kolb (1984) and Honey &Mumford (2002). He also draws attention to the importance of context in the form of the organisational setting for much learning (see Argyris and Schon (1974) on The Learning Organisation and Marsick and Watkins (2018). on team learning Arguably there is much of significance in this early work that can provide a starting point to evaluating the extent and direction of learning.

Two other frameworks cited in the literature take a rather more unrestricted approach, relying more on a naturalistic, non-interventionist methodology for data gathering. They tend to start from a hypothesis of what has occurred and seek to prove or disprove this stance. Realist Evaluation associated with Pawson and Tilley (1997) explores what works, for whom, in what circumstances, in what respects, to what extent, and why. Abma and Stake (2014) rely on natural conversations and particular observations above data generated from responses to instruments devised by the researcher; they contend that it helps to view multiple perspectives in order to see meaning and context. They are concerned less with the participant and the intervention and more with the sociological factors that influence the outcome.

Abma and Stake (2014) base their approach on the model first developed by Stake in the 1960s. They propose that any context contains "emic issues", particular cultural, internal elements and their functioning and these issues can be identified via discourse with stakeholders. "Etic issues" are brought in from outside and these may include external factors such as policy, finance. The approach seeks to interpret rather than explain and examines multiple perspectives to arrive at a holistic picture. Thus there is sensitivity to surrounding forces, and understanding of multiple sequenced, contextual, organic varieties of perspective. They stress that their approach is case specific although it may be possible to make generalisations from the "thick descriptions" that emerge.

It would appear that there are numerous models of evaluation approaches on offer that, with little variation in terminology, encompass the learning of the individual within a context specific environment. This begs the question of why there is a paucity of evidence of their use. Good data collection, analysis and conclusions are scarce. Notwithstanding the broad

focus on measurement, it remains demanding to identify the "relevant factors" that should populate a framework, an issue highlighted by Abma and Stake with their emic and etic issues. Little attention is given to a "before and after" assessment of change. Too often, it is observed, opportunistic data is used, with little regard to its value and relationship with the intervention being measured. Trojan et al., (2009) propose a Human Systems Dynamic approach. Buriak and Ayars (2019) use an approach which bears similarities with Statistical Process Control, but which they call Interrupted Time Series. Both are based on the concept of measuring pre-and post intervention via runs of data to identify changes. These methods allow calculation of pre- and post-intervention results, assessment of initial and sustained impacts of interventions and identification of pre-intervention baseline trends (Somers et al., 2013). The methods require the identification of relevant factors, data gathering, and charting over a period of time, resulting in the identification of change.

In summary, measuring learning is viewed as necessary but not sufficient. Clearly we need to understand participant satisfaction and degree of learning achieved. But this cannot constitute an entire evaluation approach. The problem of examining organisational context lies at the core of the evaluation conundrum. It is a fairly straightforward issue to measure and analyse pre-and post intervention learning; behaviour change is more elusive but can be identified; organisational impact, given the large number of factors involved, remains challenging.

Discussion

The most widely cited framework used within educational evaluations is Kirkpatrick's model. There has always been some dispute about whether Kirkpatrick gave us a model, a framework, or taxonomy, but the man himself said, when asked, that it didn't matter so long as it was used. Its popularity is probably due to its capacity to be intuitively understood. The proposition of four "levels" of outcome following an intervention (satisfaction, learning, behaviour change, results) has been adopted and has become the norm in evaluation projects.

While categorisation in this way can often provide a useful starting point for effective evaluation, the model itself may not be ideally suited to the evaluation of health care education. The framework provides a useful ordering device but it lacks guidance on how to address the nuances of organisational context. That is, where the Kirkpatrick model emphasizes increased confidence in newly acquired knowledge as being important, effective health care education would want evidence that this new knowledge is both learned and implemented in practice beyond a practitioner's own perception of knowledge gain or confidence. For health care education to truly support health care practice we need to be able to measure accurately the added knowledge, skills, or awareness that interventions may or may not provide, but we also need to acquire evidence of the impact on practice of the intervention.

Kirkpatrick's model gives very little guidance on how to identify this kind of evidence nor how data can be analysed to enable judgements about effectiveness. Questions of bias in selection of data, as well as questions of standards and comparators are not addressed.

Particularly problematic are questions about how to isolate and measure outcomes resulting from an intervention. To what extent can it be claimed that an intervention led to improved performance without taking into consideration all other factors that have impacted alongside.

Weaknesses in the Kirkpatrick model have long been acknowledged. Kaufman and Keller (1994) in their article *"Levels of Evaluation: Beyond Kirkpatrick"* highlight the lack of regard for context in the original Kirkpatrick work and they propose an additional sub-level dividing level 1 into 1a and 1b; they suggest that not only reaction to learning, but focus on what factors enable learning should be taken into account. In other words they explore the context within which learning happens. Equally they propose that an additional level should be added, which examines social outcomes. Alexander and Christoffersen (2006) in their article "*Total Evaluation Process*," cite Phillips by suggesting that Kirkpatrick should have an additional level measuring ROI. How this might be achieved in a service industry is not explored!

The significance of context has exercised the minds of many of those proposing evaluation frameworks, including Stufflebeam, Donabedian and Stake. For example Stufflebeam's CIPP model claims to be decision-oriented, focusing on

"systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming," Patton (1997, p.23).

The CIPP evaluation framework, it is claimed, can guide evaluations of programmes, projects, personnel, products, institutions, and evaluation systems (Stufflebeam 2001). However the manner in which context is examined concentrates on the wider view of where both the programme and the evaluation fit, with a tendency to describe rather than judge. Although this can assist in decision-making related to planning, and can enable the evaluator to identify the needs, assets, and resources of a community it is narrow in providing substantive critique. Context evaluation will clearly enable identification of the political climate that will influence the success of the programme. (Mertens and Wilson, 2012) In addition, the model encourages the assessment of programme goals using both formative and summative measures, such as environmental analysis of existing documents, programme profiling, case study interviews, and stakeholder interviews (Mertens and Wilson, 2012). The model appears to be strong on data gathering but less so on analysis and judgement.

While Stufflebeam was developing CIPP, Donabedian was proposing a similar model, focusing on the systems that underpin the quality of delivery of healthcare by examining structure, process, and outcomes. Both are interested in the process of change, but whereas Stufflebeam is evaluating educational interventions, Donabedian focuses on the delivery of quality. Both models are useful in drawing attention to the physical and political environment within which change takes place, but they fall short of assessing the extent to which this contextual background supports or inhibits the application of learning. Donabedian is concerned with the quality of care and its capacity to be judged using his framework within

the many healthcare systems found, and his model is simple and straightforward, but its focus on the system arguably limits its value.

In contrast, the model proposed by Stake is undoubtedly three dimensional, and it provides a holistic description and judgement of a learning intervention, as well as enabling comparison with others.

Stake suggests that in the first instance, the focus of an evaluation should reside with the rationale underpinning the intervention. He suggests that conversations be had with all stakeholders about their perceptions of why the intervention is being done, its purpose and intended outcomes, and what in their view would signal success or failure. It is highly likely that different stakeholders will have differing views; a commissioner may envisage the intervention as enabling organisational change, whereas a senior manager may view it as an individual development opportunity. Thus judgements will be made using different criteria and thus the rating of the intervention will vary. It is arguable that such details are not conventionally explored at the planning stage.

Stake then proposes that attention be given to what was intended to happen before, during and after the programme, and how closely intentions are matched by the actuality of what takes place. He suggests that all stakeholders be involved in discussions and that they are enabled to hear alternate views. At the same time Stake suggests making comparisons with external good practice so that judgements are informed by the wider experience of others. The model focuses on the closeness of intentions and the reality of what occurs and also the contingent steps that need to be in place at each stage of development and delivery of the intervention.

Stake's model is detailed and arguably resource intensive. It takes a case study approach, and does not claim generalisability, but lessons learned from an evaluation can inform future iterations of an intervention.

Although differences in terminology make for problems in direct comparisons, we have attempted to draw out the chief features of each framework (see table 2).

Insert table 3 about here

Each model has strengths and weaknesses and the value of each framework should be judged on the basis of the intervention under scrutiny and the focus of the evaluation.

Conclusion

This scoping review used the Arksey and O'Malley (2005) methodological framework for the credible and reflexive examination of a breadth of literature and to identify a research question to guide this review. The research question sought to identify the research addressing what is known from the existing literature about the effectiveness of a range of models/ frameworks that purport to evaluate the impact of learning programmes. On

reflection, terms such as effectiveness and impact were difficult to define and the existing body of research fails to adequately address such terms. This paper adds to the literature in that we critique and compare key features of models and frameworks which are popular within the existing literature for evaluating work-based learning programmes.

A programme of learning that includes the undertaking of a work related project offers scope for an extensive evaluation that will offer guidance on sustainability and spread. Such an evaluation will contain components that cover the breadth of knowledge and skills acquired by participants within a context that is unique and influential.

Although there is much in the academic literature that provides a grounding for thinking about evaluation approaches, there is a dichotomy between what is claimed and what actually occurs in most cases. Theoretical underpinning is weak, techniques are superficial and difficult to replicate. In this paper we have attempted to gather together evaluation techniques that can offer greatest support to structuring, undertaking, analysing and reporting on the outcomes of work place learning programmes.

It is evident from the literature that the Kirkpatrick framework is the most widely known and (it is claimed) used. But it is generally acknowledged that the framework suffers from uneven facility in use, such that the tendency is to concentrate on the first of the four levels described. Notwithstanding, the four levels (with slight variations in terminology) appear in most frameworks that seek to cover the range of elements of activity involved in such an enterprise.

Kirkpatrick can be used as a framing mechanism but requires additional input. This could include approaches and frameworks proposed by other researchers, to ensure detailed guidance for a comprehensive evaluation. The implication of this review, for those involved in developing and undertaking the evaluation of work-based learning, is the need to adopt a holistic model which embraces learning, process, context, and systems. This will enable judgements to be made about learning and skills acquired, and their use in a setting that is unique and challenging. This expands the extensively used Kirkpatrick framework to encourage coverage of all the levels. This will enable a comprehensive picture of the success of a programme can be adjudged.

Selection of a complement to Kirkpatrick will depend on the intervention under scrutiny, and arguably Stake, Stufflebeam and Donabedian each have something to offer. For example, the Kirkpatrick and the Stake models together provide a comprehensive and multi – dimensional approach to examining the development, delivery and outcomes of an intervention, capturing all elements, all relevant data, and providing a clear strategy for analysis of findings, thus offering a comparable and credible evaluation pathway. Equally, Kirkpatrick and Stufflebeam together offer scrutiny of learning within a systems focused dynamic. Kirkpatrick and Donabedian offer close analysis of outcomes with particular reference to the quality of care. The choice of approach to evaluation should start by agreeing purpose and rationale of the evaluation itself.

Evaluators and academics have a duty to provide sufficient details of the framework(s) employed to enable others to assess whether a hybrid approach is required. Providing detailed cases of evaluations will further enhance the quality of work-based learning and the learning associated with evaluation. Many of the papers originally selected for inclusion in this review were rejected due to limited information about the 'what' (framework(s)) and the 'how' (was the evaluation undertaken) and any adaptations made. This paper compares the attributes of four key models/frameworks which we hope will assist evaluators when designing their evaluations.

As with all scoping reviews, this research is limited due to the search terms and databases used; it is possible that using alternative words would have identified additional literature. Likewise, the parameters applied to the searches made to limit research (e.g. between 1985 – 2019, English language papers) may have inadvertently excluded earlier research or international research. Therefore, it could be argued that use of a systematic review methodology would be more appropriate. However, this review offers insight and critique of existing models and frameworks used to evaluate work-based learning. Here we have compared the key features of the frameworks we have reviewed which we hope will help to clarify for future application.

Although there is much in the academic literature that provides a grounding for thinking about evaluation approaches, there is a dichotomy between what is claimed and what actually occurs in most cases. Future research needs to address the key issues raised in this review: theoretical underpinning is weak; techniques are superficial and difficult to replicate. In this paper we have attempted to gather together evaluation techniques that can offer greatest support to structuring, undertaking, analysing and reporting on the outcomes of workplace learning programmes.

References

Abma, T. andStake, R. (2014), 'Science of the Particular: An Advocacy of Naturalistic Case Study in Health Research', *Qualitative Health Research*, Vol. 24 No. 8, pp.1150-116.

Alexander, M. and Christoffersen, J. (2006), 'The Total Evaluation Process', *Performance Improvement*, Vol. 45 No. 7, pp.23- 28.

Argyris, C. and Schon, D. (1974), *Theory in Practice: Increasing Professional Effectiveness,* Jossey Bass, San Francisco.

Arksey, H. and O'Malley, L. (2005), 'Scoping Studies: Towards a Methodological Framework', *International Journal of Social Research Methodology*, Vol. 8 No. 1, pp.19-32.

Aziz, S. Mahmood, M. and Rehman, Z. (2018), 'Implementation of CIPP Model for Quality Education at School Level: A Case Study', *Journal of Education and Education Development*, Vol. 5 No. 1, pp.57–64.

Baillie, L. Bromley, B. Walker, M. Jones, R, and Mhlanga, F. (2014),'Implementing service improvement projects within pre-registration nursing education: A multi-method case study evaluation', *Nurse Education in Practice*, Vol.14 No. 1, pp.62-68.

Bentley K. (2015), 'The national diabetes education program evaluation framework: How to design an evaluation of a multifaceted public health education program', *Journal of Social Work Education*, Vol.49, pp.78-84.

Billett, S. Smith, R. and Barker, M.(2005), 'Understanding work, learning and the remaking of cultural practices', *Studies in Continuing Education*, Vol. 27 No. 3, pp.219–237.

Billett, S. Higgs, J. Barnett, R. Hutchings, M. and Trede, F.(eds) (2012), *Practice Based Education*. Springer, Dordrecht.

Buriak, S. and Ayars, C. (2019), 'Evaluation of a Drug and Alcohol Safety Education Program using Interrupted Time Series and the Kirkpatrick Framework', *Evaluation and Program Planning*, Vol.73, pp.36-65.

Campbell, K. Taylor, V. and Douglas, S. (2017), 'Effectiveness of Online Cancer Education for Nurses and Allied Health Professionals: a Systematic Review using Kirkpatrick's Evaluation Framework', *Journal of Cancer Education*, Vol. 34 No. 2, pp.339-356.

Carr, G. Loucks, D. and Blöschl, G. (2014), 'Introducing a new learning and teaching evaluation planning framework for small internally funded projects in higher education', *Research Policy*, Vol. 47, pp.57-67.

Chmiel, A. Shaha, M. and Schneider, D. (2017), 'Development and Evaluation of an International Service Learning Program for Nursing Students', *Nurse Education Today*, Vol. 02606917.

Clark, M. Hutchison, C. and Lockyer, J. (2010), 'Musculoskeletal education: a curriculum evaluation at one university', *BMC Medical Education*, Vol. 10, p.93.

Craig, P. Barnard, A. Glasgow, N. and May, E. (2014), 'Evaluating the Health "Hubs and Spokes" interprofessional placements in rural New South Wales, Australia', *Journal of Allied Health*, Vol. 43 No. 3, pp.176-83.

Curado, C. and Teixeira, S. (2014), 'Training evaluation levels and ROI: the case of a small logistics company', *European Journal of Training and Development*, Vol. 38 No. 9, p.845-871.

Curtin, A. Martins, D. Schwartz-Barcott, D. DiMaria, L. Ogando, B. and Milagros, S. (2013), 'Evaluation of Technology-Enhanced Learning Programs for Health Care Professionals: Systematic Review'. *Public Health Nursing*, 07371209.

Derven, M. (2012), 'Building a Strategic Approach to Learning', *Evaluation Training & Development*, Vol. 66 No. 11 pp.54-57.

Dewey, J. (1963), *Experience and Education*, 4th edition Collier Books, New York, NY

Donabedian, A. (1981), 'Criteria, Norms and Standards of Quality: What do they Mean?' *American Journal of Public Health*, Vol.71 No. 4, pp.46-56.

Donabedian, A. (2005), 'Evaluating the Quality of Medical Care', *Milbank Quarterly*, Vol. 83 No. 4, pp.28-42.

Engestrom, Y. (2016), *Studies in Expansive Learning*, Cambridge University Press, England.

Eraut (2002) Do the concepts of "learning community" and "community of practice" provide added value? *ESRC seminar* February 2005

Freeth, D. Reeves, S. Zwarenstein, M. Goldman, J. Barr, H. Koppel, I. and Hammick M. (2010), 'The Effectiveness of Interprofessional Education: Key Findings from a New Systematic Review', *Journal of Interprofessional Care*, Vol. 24 No. 3, pp.230-241.

Fu, A. Peterson, L. Kannan, A. Shavelson, R. and Kurpius, A. (2012), 'The 2011 Program Evaluation Standards: A framework for quality in medical education programme evaluations' *Visitor Studies*, Vol. 18 pp.67-82.

Gallivan, J. Greenberg, R. and Brown, C. (2008), 'The National Diabetes Education Program evaluation framework: how to design an evaluation of a multifaceted public health education program', *Preventing Chronic Disease*, Vol. 5 No. 4, pp.A134.

Gallivan, J. Greenberg, R. and Brown C. (2012), 'Academic microsystems: Adapting clinical microsystems as an evaluation framework for community-based nursing education', *Preventing Chronic Disease*, Vol. 5, pp.35-43.

Henry, C. (2015), 'Entrepreneurship education evaluation: revisiting Storey to hunt for the heffalump', *Education + Training*, Vol. 57 No. 8/9, pp.816-833.

Horwich, P. (2004), 'A Use Theory of Meaning', *Philosophy and Phenomenological Research*, Vol XV111 No 2, pp. 351-372.

Huber, E. (2016), 'A Framework for Summative Evaluation in Informal Science Education', *Journal of University Teaching and Learning Practice*, Vol. 14, pp165-172.

Jarvis, P. (1987), *Adult Learning in the Social Context*, Croom Helm, London.

Johnston, S. Coyer, F. and Nash, R. (2018), 'Kirkpatrick Evaluation of Simulation and Debriefing in Health Care Education: a Systematic Review', *The Journal of Nursing Education*, Vol. 57 No. 7, pp.393-398.

Kaufman, R. and Keller, J. (1994), 'Levels of Evaluation Beyond Kirkpatrick', *Human Resource Development Quarterly*, Vol. 5 No. 4, pp.65-78.

Kirkpatrick, D. and Kirkpatrick, J. (2006), *Evaluating Training Programs*, Third Edition. Berrett-Koehler, San Francisco.

Knowles, M. (1980), *The modern practice of adult education: from pedagogy to adragogy*, Associated Press, Wilton Connecticut.

Kolb, D. (1984), *Experiential Learning: Experience as a Source of Learning and Development*, Prentice Hall, New Jersey.

Koyama, M. Holzemer, W. Kaharu, C. Watanabe, M. Yoshii, Y. and Otawa, K. (1996), 'Assessment of a continuing education evaluation framework', *Journal of Continuing Education in Nursing*, Vol. 27 No. 3, pp.115-119.

Larkins, S. Preston, R. Matte, M. Lindemann, I. Samson, R. Tandinco, F. Buso, D. Ross, S. Pálsdóttir, B. and Neusy, A. (2013), Measuring social accountability in health professional Education: Development and international pilot testing of an evaluation framework, *Medical Teacher*, Vol. 35 No. 1, pp.32-45.

Lechner, S. Lechner, K. and Thomas, G. (1999), 'Evaluation of a computer-aided learning program in removable partial denture framework designing', *Journal of Prosthodontics*, Vol. 8 No. 2, pp.100-105.

Lewin, K. (1973), *Resolving Social Conflicts*, Tavistock Publishers, London.

Marsick, V. Watkins, K. Wofford, M. Ellinger, A. (2018), 'The Evolving Marsick and Watkins 1990) Theory of Informal and Incidental Learning', *New Directions in Adult and Continuing Education*, Vol. 159, pp.21-36.

Mcmahon, J. and Cullinan, V. (2014), 'Education programmes for young children with Autism Spectrum Disorder: An Evaluation Framework', *Research in Developmental Disabilities*, Vol. 35 No. 12, pp.3689-3697.

Mertens, D. and Wilson, A. (2012). *Program Evaluation Theory and Practice*, Guilford Press, New York.

Mezirow, J. (1997), 'Transformative Learning: Theory to Practice', *New Directions for Adult and Continuing Education*. Vol. 74, pp.5-12.

Mumford, A. (2002), 'Achieving Results through Action Learning', *Industrial and Commercial Training*, Vol. 34 No. 1, pp.28-30.

Nicoll, P. MacRury, S. van Woerden, H. and Smyth, K. (2016), 'Learning how programs achieve their impact: embedding theory-driven process evaluation and other program learning mechanisms in alive & thrive', *Journal of Medical Internet Research*, Vol. 4, pp.35-54.

Patton, M. (1997), *Utilization-focused Evaluation: the new century text* (3rd edition), Sage Publications, Thousand Oaks CA.

Pawson, R. and Tilley, N. (1997), *Realistic Evaluation*, Sage Publications Ltd. London.

Peterson, J. Pearce, P. Ferguson, L. and Langford, C. (2017), 'Understanding scoping reviews: Definiion, purpose and process', *Journal of the American Association of Nurse Practitioners*, Vol. 29, pp.12–16.

Praslova, L. (2010), 'Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in Higher Education', *Educational Assessment, Evaluation and Accountability*, Vol. 22 No. 3, pp.215–22.

Rawat, R. Nguyen, P. Ali, D. and Saha K. (2013), 'Evaluation of a drug and alcohol safety education program in aviation using interrupted time series and the Kirkpatrick framework', *Food & Nutrition*, Bulletin Supplement 2.

Revans, R. (1978), *ABC of Action Learning*, Lemos and Crane. London.

Ruhe, V. and Boudreau, J.D. (2014), 'Toward an evaluation framework for doctoral education in social work: A 10-year retrospective of one PhD program's assessment experiences', *Journal of Evaluation in Clinical Practice*, Vol. 19, pp.74-84.

Singh, M. (2004), 'Evaluation framework for nursing education programs: application of the CIPP model', *International Journal of Nurse Education Scholarship*, Vol.1, Article13.

Somers, A. Robays, H. DePaepe, P. Van Maele, G. Perehudoff, K. and Petrovic, M. (2013), 'Evaluation of Clinical Pharmacist Recommendations in the Geriatric Ward of a Belgian University Hospital'*, Clinical Interventions in Aging*, Vol. 8, pp.703-709.

Stake, R. (1967), *The Countenance of Educational Evaluation*, available at https://pdfs.semanticscholar.org/b07e/5b61cde550bfb0b64e895674a236c9003335.pdf [Accessed on 10.10. 20].

Stern, E. Jagosh, J. and Tilley, N. (2016), 'Realist Evaluation at 25: Cumulating Knowledge, Advancing Debates and Innovating Methods', *Evaluation*, Vol. 22 No. 3, pp.267-269.

Stufflebeam, D. (2001), 'The Metaevaluation Imperative', *American Journal of Evaluation*, Vol. 22 No. 2, pp.46-62.

Tadesse, T. Mengistu, S. and Gorfu, Y. (2016), 'Using research-based evaluation to inform changes in the development of undergraduate sports science education in Ethiopia', *Journal of Hospitality, Leisure, Sport & Tourism Education*, Vol. 18, pp.42-50.

Thies, K. and Ayers, L. (2007), 'Academic microsystems: adapting Clinical Microsystems as an evaluation framework for community-based nursing education', *Journal of Nursing Education*, Vol. 46 No. 7, pp. 325-329.

Thistlethwaite, J. Kumar, K. Saunders, R. and Carr, S. (2015), 'An Exploratory Review of Pre-Qualifying IPE Evaluations', *Journal of Interprofessional Care*, Vol. 29 No. 4, pp. 67-82.

Trojan, L. Suter, E. Arthur, N. and Taylor, E. (2009). 'Evaluation Framework for a Multi-site Practice Based Interprofessional Education Intervention', *Journal of Interprofessional Care*, Vol. 23 No. 4, pp.380-389.

Unwin, L. and Fuller, A. (2005), *Learning as work: teaching and learning processes in the Contemporary work organisation*, University of Leicester, Leicester.

Vygotsky, L. (1978), *Mind in Society*, Harvard Business Press, Massachusetts.

Wenger, E. (1998), *Communities of Practice: Learning as a social system*, available at http://www.co-i-l.com/coil/knowledge-garden/coplss.shtml (accessed on 10.02.20).