Evaluating leadership development in a changing world? Alternative models and approaches for healthcare organisations

Dr Paul Joseph-Richard & Professor Janet McCray

Published online: 24 Feb 2022.

Submit your article to this journal ⤤

Article views: 490

View related articles ⤤

View Crossmark data ⤤

Evaluating leadership development in a changing world? Alternative models and approaches for healthcare organisations

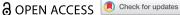Dr Paul Joseph-Richard [a] and Professor Janet McCray [b]

aLecturer in HRM, Ulster University Business School, Ulster University, BT37 0QB Jordanstown, UNITED KINGDOM; bProfessor of Social Care and Workforce Development, Department of Childhood, Social Work and Social Care, University of Chichester, College Lane, Chichester, West Sussex

ABSTRACT

Internationally, healthcare is undergoing a major reconfiguration in a post-pandemic world. To make sense of this change and deliver an integrated provision of care, which improve both patient outcomes and satisfaction for key stakeholders, healthcare leaders must develop an insight into the context in which healthcare is delivered, and leadership is enacted. Formal leadership development programmes (LDPs) are widely used for developing leaders and leadership in healthcare organizations. However, there is a paucity of rigorous evaluations of LDPs. Existing evaluations often focus on individual-level outcomes, with limited attention to long-term outcomes that might emerge across team and organizational levels. Specifically, evaluation models that have been closely associated with or rely heavily on qualitative methods are seldom used in LDP evaluations, despite their relevance for capturing unanticipated outcomes, investigating learning impact over time, and studying collective outcomes at multiple levels. The purpose of this paper is to review the potential of qualitative models and approaches in healthcare leadership development evaluation. This scoping review identifies seventeen evaluation models and approaches. Findings indicate that the incorporation of qualitative and participatory elements in evaluation designs could offer a richer demonstration and context-specific explanations of programme impact in healthcare contexts.

Introduction

There is no longer a conventional, definitive context for healthcare services, as delivery systems, models, and professional practice continue to change (Kings Fund 2021) in the wake of the global pandemic of 2020. Future challenges to be faced include the sustainability of health and social care systems and the support of professional staff to enable them to practice (World Health Organization (WHO), 2020). Globally, the WHO draws attention to the need for a focus on innovation created by digital transformations, calls for service improvement, and a further roll out of accountable or integrated care by reorienting health systems towards a collaborative primary care approach with team-

based care (WHO, 2020). To achieve this, there is a need for patient involvement and localized solutions at the core, with the WHO recommending that all nations embed stakeholder engagement in their healthcare and workforce planning strategies to address the complex issues faced by healthcare organizations.

Global healthcare systems are highly complex institutions. These systems provide services to a heterogeneous population where individuals with complex mental and physical conditions need inputs from professionals, services, and systems that are inter-dependent yet often function separately (Aveling, Parker, and Dixon-Woods 2016). Many systems struggle to balance the operational aspects of compassionate care giving and leadership in their hospitals (West, 2021), with the necessity to answer questions related to effectiveness and efficiency of their services, quality of care, patient expecta-tions, and 'what works' agenda (Long 2006). In economically developed countries, the drive for patient safety and care efficiency has created moves towards the standardization of care processes. Despite support in principle, tensions between managers and clinical experts may exist, as professional judgement is viewed as being eroded, and replaced by unquestioned rule following (Martin et al. 2017). Adding further complexity is how hospitals and services operate, and within these hospitals, how different professional groups respond to leaders and leadership (Andersson 2015). For whilst global integrated care reports, and policy-makers highlight the importance of collaboration in organiza-tions and those who work in them, continents, and countries have different cultural values, spending priorities, and funding streams within healthcare structures. For exam-ple, the continents of North America and Canada are developing collaboration based on a matrix leadership structure (Okpala 2020). However, complex bureaucracy makes practice change challenging (Kuluski & Reid, 2020). In South America, a top-down leadership style is predominant with employees avoiding conflict and not tending to speak out (Maddox and Replogle 2019). Hierarchical structures in hospitals may be as prevalent as in China, which is undergoing major healthcare reform (Yang et al. 2020), and in Africa, where communalism and non-individualism based on the African values of Ubuntu (Olano 2015) are viewed as essential for twenty-first century leadership in healthcare services. Raju (2021) notes that in Northern India there is a lack of preparation for clinicians to undertake any form of leadership role and a culture lacking in trust across professionals is prevalent. Within Europe, leadership in healthcare is also often categorized as hierarchical and transactional (Sola et al, 2016); however, there is evidence that this is changing. In Germany, consensus amongst staff teams is considered impor-tant and research in acute hospitals correlates transformational leadership behaviour and reporting of critical safety incidents (Hillen, 2017). Within healthcare organizations, HRD planning solutions for transforming services may assume there is a desire within different professional groups and managers to come together to make integrated care happen. The reality may be more nuanced because the day-to-day pressures of providing healthcare take priority. All these factors contribute to the different ways leaders and leadership are defined, leadership structures are institutionalized, and goals of leadership development are agreed, in global healthcare systems. Yet, there is a need to focus on ensuring care delivery and building support for collaboration (Nuno-Solimis (2017), a key element of collaboration and organizing care being leadership, and subsequently for HRD practitioners' investment in the best fit leadership development (Sfantou et al. 2017).

To develop leaders and leadership, global healthcare organizations continue to invest in formal LDPs (Turner 2019; Ho 2016). However, very few organizations believe their LDPs are highly effective (Schwartz, Bersin, and Pelster 2014), calling into question the effectiveness of current LDPs (Lacerenza et al. 2017). As healthcare organizations are being challenged to demonstrate the impact of their LDPs, they recognize that they cannot rely solely on traditional individualistic leadership development models and approaches, to develop, for example, learners who come from diverse cultural and societal contexts and who are patient focussed (McCray, Temple and McGregor, 2021). In fact, the LDPs that seek to develop individuals' personal development are under scrutiny, as views of the 'given' competencies and characteristics of leaders in interdependent healthcare systems are reviewed. As Ham, Berwick, and Dixon (2016) note, leaders require boundary and hierarchy-spanning skills to negotiate systems and work across care settings with other professionals, patients, and other local stakeholders. These skills are needed to drive innovation and to contribute to more equal partnerships in service improvement. Moreover, effective leaders enact their leadership through socially and situationally constructed collaboration and inter-professional partnerships (McCauley and Palus 2020). Bate et al. (2014) advise that moving away from the notion of the healthcare context as a 'fixed entity', which is capturing the most predictable of outcomes, to one which can also highlight the interactions between stakeholders, i.e., patients and staff on the care delivery pathway, enables a more holistic explanation for leadership actions. Here, the context is integral rather than something that is unchanging throughout the social transformation or other processes (Pettigrew, Woodman, and Cameron 2001). Thus, as systems, alliances, and alignments in care delivery are changing, LDPs have begun to change their curricula in response to these changes. Human Resource Development (HRD) scholars, practitioners, and commissioners are therefore, moving away from programmes that are built upon the universal and individualistic models of leadership (Leach et al. 2021; Ford 2015; West et al. 2015, 2021; Edmonstone 2013a, 2013b, 2011) as these models are not wholly sufficient to meet the more nuanced leadership learning needs, in healthcare contexts.

As traditional LDP programmes are being re-designed, the evaluation models and approaches that are applied for LDP evaluations may also need to change. Evaluation is understood as 'a process of determining the merit, worth or value of something, or the product of that process' (Scriven 1991, 139). As many qualitative evaluation models and approaches have remained under-utilized, we undertake a scoping review to (1) consider the strengths and weaknesses of the dominant, traditional evaluation models for application in present-day healthcare organizations; (2) offer additional models of leadership development evaluation for application in LDP healthcare contexts by outlining their potential contribution during radical change, and (3) suggest how these alternative models can offer complementary tools to capture the impact of leadership development programmes in healthcare contexts. We argue that the intentional use of qualitative evaluation models and approaches, along with the traditional evaluation models, such as the one proposed by Kirkpatrick (1996) and others, may help HRD practitioners and other leadership developers to prove programme worth and improve programmes. Our review makes an important contribution to healthcare LDP evaluation literature by shining light on a range of qualitative evaluation models and approaches that have

been previously underutilized but merit a renewed attention from HRD scholars and practitioners and offers an improved understanding of their strengths and weaknesses, using Mabey's (2013) leadership development discourse framework.

Theoretical background: evaluation models/approaches

In the programme evaluation literature, the terms 'model' and 'approach' have generally been used interchangeably (Bennett 2003). While the term 'approach' is used to cover an eclectic set of good evaluation practices, the term 'model' is often used for labelling 'idealized ... views for conducting programme evaluations according to their authors' beliefs and experiences' (Stufflebeam and Shinkfield 2007, 135). Very often, these models and approaches offer a set of recognizable ways to design evaluations and implement them in specific contexts, therefore, empowering the evaluators to conduct more meaningful evaluations. To determine the merit or worth of a programme, most evaluators tend to be concerned with the questions of its impact (e.g., what the outcomes of a particular programme in question are). The approach of many practitioners is to focus on both quantitative and qualitative factors (Finney and Jefkins 2009) that could help them demonstrate programme value. But in the context of an actual project, and with the pressure to demonstrate project investment returns, evaluators may be forced to adopt a 'functionalist-mindset' (Mabey 2013) that makes them stay focus only on a limited set of 'see-able' outcomes; as they direct their attention towards demonstrable short-term evidence, they tend to ignore both the context and the other possible outcomes that emerge with a passage of time. In what follows, we examine why evaluating leadership development programmes in healthcare contexts is challenging.

Evaluating leadership programmes

Identifying, measuring, and demonstrating the impact of LDPs is challenging because of the inherent complexities enshrined in programme design and delivery. Hartley, Martin, and Benington (2008) argue that 'in order for evaluation to occur with any degree of robustness, there is a need for a reasonably clear specification of what forms the basis of the leadership development, leadership, and organizational performance' (170). In practice, however, there can be ambiguity in what is being developed, how and why, in these programmes (Day 2011), and most LDPs are not always guided by any leadership theory (Avolio et al. 2009). There is also a lack of empirical support for the effectiveness of the developmental methods used in LDPs (Burgoyne, Hirsh, and Williams 2004), as many LDPs fail to link learning experiences with the challenges of delivering value at work. Crucially, there is also a tendency among evaluators to overlook the influence of context on leadership learning, reflection, and application (Edwards and Turnbull 2013a, 2013b). Yet, the national, regional, local, and within institution cultural contexts play a significant role in the ways leadership is understood, enacted and developed. In many cases, the programme context, the power-relations, and the dominant cultural values determine how leader identity is developed (Gagnon and Collinson 2014), and how LDPs' appropriateness and relevance are judged.

Evaluating LDPs becomes even more challenging when leader development of individuals and leadership development of the collectives are not well understood. Mabey (2013), when presenting an insightful framework of discourses on leadership development, argues that within the complex and volatile contexts such as in healthcare, a solely functionalist mindset towards evaluation is problematic. A functionalist perspective refers to a fixation with (a) enhancing the under-developed qualities of individual leaders through formal LDPs, as if they are in perceptual need for trainer-centred skill development events and the developed individuals will be personally capable of lifting others' performance and of transforming complex healthcare organizations; and (b) that evaluations can faithfully and robustly capture the knowledge, skills, and attitude gains that are experienced by LDP attendees, as if these outcomes are the only critical ones for the leaders that are making hospitals responsive and efficient. Although an evaluation based on the functionalist mindset may be useful, this perspective emerges from a narrow view of individualistic leadership, that has been (mistakenly) assumed to emerge in a social or cultural vacuum, and from the view that ignores the larger cultural, economic, institutional, and societal pressures that shape leaders and leadership learning. However, the current healthcare context sees this position changing. Mabey (2013) proposes that we should complement our understanding of leadership from the interpretive, dialogic, and critical perspectives on leadership and leadership development. Since Mabey's (2013) discourse framework has been recognized to have the potential to enhance our understanding of leadership development, and to make us 'better informed and critical' learners of leadership development (Carroll 2019, 127), we use his framework to categorize our findings.

The dominance of a functionalist leadership development discourse

In a challenging context, Kirkpatrick's taxonomy (See Table 1A) may be perceived as an easy-to-use standard for demonstrating the impact, of complex programmes. Owing to its conceptual simplicity (Russ-Eft and Preskill 2009), prescriptive appeal and high face validity (Arthur et al. 2003), this taxonomy has become the most widely used framework for not only supervisory training programmes (for which it was originally intended) but also for all types of learning and development programmes (Hoole and Martineau 2014; Collins and Denyer 2008), including LDPs (Ely et al. 2010; McLean and Moss 2003), particularly in healthcare organizations (King and Nesbit 2015).

Despite its popularity and its significant impact on evaluation practices in healthcare, researchers (such as Holton 1996; Bates 2004; Anderson 2010) have identified several limitations of this taxonomy (and of other taxonomy-based models proposed by Swanson and Sleezer, 1987, for example). Critics are concerned with the absence of a causal link that Kirkpatrick presumed to exist between the levels (Alliger et al. 1997); they argue that the term 'learning' was conceptualized too narrow to include only knowledge, skills, and attitudes while ignoring the more complex, contextual learning that is enriched by ongoing reflection (de Dea Roglio and Light 2009), and changes in mindsets (Kennedy, Carroll, and Francoeur 2013). Also, scholars are concerned with problematic assumptions that underpin this taxonomy: that all participants are the same, that every learner will complete the programme and will transfer the learning, and that all this transferred learning can be measured as observable behaviour (Ford and Sinha 2008;

Table 1A. Examples of models* using a 'Taxonomy-of-outcomes', based on a functionalist perspective.

Evaluation model	Approach	Underlying Assumption(s)	How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist advocate
Kirkpatrick's Taxonomy	A taxonomy that specifies four sets of outcomes that should be considered as 'targets' of evaluation. These outcomes are Participant Reactions, Learning, Behaviour and Results.	Individuals are in need of development. Trainer-focused, classroom-based interventions could improve deficiencies in learners' knowledge, skills, and attitudes. Four outcomes, if captured in full, can demonstrate the value of an intervention.	• Collect learner reactions at the end of a programme or immediately after the programme. • Measure learning outcomes using pre- and post-tests. • Collect evidence for application of learning (e.g. change in behaviour) • Collect evidence for system-level outputs and outcomes such as reduced turnover and increased satisfaction.	Assumed suitable for LDPs in all kinds of healthcare contexts (e.g., leader development in acute hospitals, ambulatory surgery centres, children's' hospitals, small clinics, community settings, independent hospitals, teaching and non-teaching hospitals, Trust hospitals).	• Widely used model. • Easy to use. • Can be used in part (i.e. by collecting evidence for one or two outcomes, a report can be prepared, although such partial effort is neither recommended nor useful).	• Represents functionalist perspective on leadership and leadership development. • Ignores embedded nature of leadership. • Ignores the importance of context and culture that influence the practice of leadership in healthcare settings.	Kirkpatrick (1996)
Swanson & Sleezer Taxonomy	A taxonomy that specifies three outcomes (Satisfaction of trainees and that of their supervisors, Learning, Performance) as targets of an evaluation, along with a cost-benefit analysis of a training programme.		• Present a coherent report based on all the evidence that supports the four outcomes. • Swanson provides tools for planning evaluations, gathering data and reporting the findings, and these tools are readily applied by evaluators.		• High face validity • No need for special training.		Swanson & Sleezer (1987)

*There are other less well-known models that use taxonomy-of-outcomes (e.g. Hamblin 1974; Kearns and Miller, 1997) that are reviewed in Tamkin, Yarnall, and Kerrin (2002) but are not included here.

Table 1B. Other methods/approaches aligned with functionalist perspectives on leadership and leadership development.

Evaluation model	Approach	Underlying Assumption (s)	How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/ advocate
Return-on-Investment (ROI) Model	A process model that promotes a step-by-step procedure for developing ROI of LDPs. Advocates collecting evidence for five outcomes.	• Isolation of programme effects from other factors in order to show that monetary payoff of LDPs is possible. • Converting benefits into monetary values is possible.	• Have an evaluation plan. Consider evaluation purpose, data collection tools, timing, and levels. • Collect quantitative and qualitative evidence for various outcomes: Reaction and planned action (Level 1); Learning (Level 2); Job application (Level 3); Business Results (Level 4); and ROI (Level 5). • Isolate the effects of the programme. • Identify intangible benefits • Convert data into monetary values, using recommended strategies. • Calculate ROI (%) = Net programme benefits/ programme costs × 100.	More suitable for LDPs that has a start date and an end date, and for those that have observable outcomes that are predetermined.	• May promote accountability. • Some stakeholders could see ROI as more convincing. • Could be used for marketing and learner recruitment processes.	• ROI model evaluates the impact at 'one point' in time, and ignores emerging outcomes, over time. • Some might consider that ROI could not be adequately measured in health contexts. • Might require specialist skills. • Published healthcare LDP ROI studies overly-rely on self-reported outcomes (Jeyaraman et al. 2018)	Phillips and Phillips 2007

(Continued)

Table 1B. (Continued).

Evaluation model	Approach	Underlying Assumption(s)	How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/ advocate
Theory based evaluation	A explicit theory (an underlying logic) of how a programme is understood to produce certain outputs and outcomes, guide the evaluation process. This theory specifies programme context, its connections between actual inputs, processes and activities, expected short-term outputs and long-term outcomes.	If a programme is based on a sound theory, then this theory can help structure and guide the evaluation. The evaluator examines whether or not the programme worked as theorized.	• Consult stakeholders and create a 'logic model' that attempts to explain how inputs are thought to produce outputs and outcomes. • Examine inputs, processes, outputs, and outcomes using multiple methods. • Report conclusions on the basis of, 'if this theory is correct, then these outputs and outcomes could be expected (Fitz-Gibbon and Morris 1987).	Most suitable if a LDP is based on a particular leadership development model OR a particular leadership theory. Since leadership development is a lengthy process that aims at distant and intangible outcomes, these theory-based evaluations can be useful to produce meaningful conclusions about its operations and possible outcomes. Most suitable when professional evaluators, knowledgeable about logic models and programme theory are available to lead evaluations.	Programme theory provides clarity and focus for evaluation. For Stame (2004) these evaluations can help build capacities in public sector and educate the public to have a better understanding of the political context in which programmes create impact. When it is not possible to do experimental evaluations, these theory-based evaluations can provide 'sufficiently compelling' evidence for causality (Weiss 1995).	• Birckmayer and Weiss (2000) report that in the published theory-based evaluations, the program theories contained unnecessary programme components, and in all the nine studies they reviewed, the original theory was not fully accurate. This explains that 'doing it right is usually not feasible, and failed or misrepresented attempts can be highly counter-productive' (Stufflebeam and Shinkfield 2007, 187). • There may be differences in what staff say they do (i.e., espoused theory) and what really happens in healthcare context (i.e., theory-in-use). When espoused theory does not match theory-in-use, then this difference could make an evaluation more complex. • Evaluators may still need other methods to determine how much the theorized results are actually produced (Funnell and Rogers 2011)	Weiss 1995

(Continued)

Table 1B. (Continued).

Evaluation model	Approach	Underlying Assumption(s)	How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/ advocate
Context, Inputs, Process, Products (CIPP Model)	A comprehensive model that specifies what to evaluate in LDP evaluation, namely programme context, its inputs and implementation process, and products (outputs and outcomes). It specifies what elements need to be assessed in each of the four programme aspects.	A system perspective is needed to generate useful managerial information.	Context: Gather information on needs, issues, opportunities, and resources in order to determine goals, priorities and judging outcomes. Inputs: Examine what goes into the programme (resources and approaches used to achieve the objectives) Process: Examine how the programme is implemented so that emergence/absence of outcomes can be explained accurately. Product: Identify outputs and outcomes so that evaluative judgement can be made in the light of all the four processes.	Suitable for LDPs that develop collective leadership at system levels (for example, in Academic Health Centers in the USA, where evaluation beyond participant satisfaction is uncommon, Lucas et al. 2018)	A more detailed and a comprehensive model that specifies what must be evaluated in a programme.	Relatively more emphasis is placed on providing information to programme managers.	Stufflebeam (1983)
CAPIRO Model Context, Administration, Process, Inputs, Reactions and outcomes	A comprehensive, accountability oriented, system based model of evaluation.		• In addition to the above, it includes evaluation of programme administration and learners' reactions.	Suitable for LDPs that develop collective leadership at system levels (for example, in Academic Health Centers in the USA, where evaluation beyond participant satisfaction is uncommon, Lucas et al. 2018).	It specifies six different types of programme elements to be included in evaluation.	Relatively more emphasis on accountability and provision of managerial information.	Easterby-Smith (1994)

Blume et al. 2010). This taxonomy also pays less attention to the many factors including the content of LDP, attendance policy, and duration of a programme that influence the effectiveness of LDPs (Lacerensa et al. 2017). In addition, it is difficult to determine how these functionalist models can take into account many of the complexities of leadership learners' behaviour and the dynamics in the radical, relational, change-landscape of post-pandemic healthcare contexts.

Similarly, scholars who adopt the functionalist perspective (e.g., Avolio, Avey, and Quisenberry 2010; Phillips and Phillips 2007) have also promoted Return-on-Investment (ROI) evaluation models, to determine the value of LDPs, although these ROIs are 'notoriously difficult to evaluate with any tangibility' (Carroll 2019, 128). A review of 138 ROI evaluations of healthcare LDPs concludes that 'the improved outcomes/ROI indicators and metrics' associated with LDPs found in majority of these studies are 'self-reported' (Jeyaraman et al. 2018), and 'the research designs varied quite widely' (Ibid, p. 87) that the authors could not assess the quality across studies. They call for more evidence-based approaches to assess the ROI of LDPs in healthcare. Some others adopting the same functionalist perspective towards LDP (Weiss 1995; Watkins, Lysø, and de Marrais 2011) have recommended a theory-based approach to evaluation. For them, a theory (or an underlying logic) of how a programme is understood to produce certain outputs and outcomes guides the evaluation process. They attempt to construct a theory that specifies programme context, actual inputs, processes, and short-term outputs while illustrating key linkages with expected long-term outcomes and impact. System-based models, such as Context, Inputs, Process, and Products (CIPP model) (Stufflebeam 1983); Context, Administration, Process, Inputs, Reactions, and Outcomes (CAPIRO model) (Easterby-Smyth 1994), also hold an instrumentalist view of LDPs. These models (see Table 1B), with predominant managerialist orientations, assume a deficit model of leadership development in individuals, who will eventually change organizational systems for the better, because of their attendance at a formal LDP. Although system-based evaluation models emphasize the role of context relatively stronger than the taxonomy-based models, they too are based on the functionalist assumption that organization-sponsored, centrally regulated formal LDPs could produce heroic leaders who could transform systems single-handedly.

Historically, LDPs have over relied on developing skills of individual leaders, whilst ignoring leadership structures and other factors, such as the tensions and power issues faced by individuals and teams (Stacey 2012, 62–65); therefore, evaluation models and approaches that are based on the functionalist view of leadership development may also tend to ignore programme context and lead to inadequate organizational learning about the LDP. In this light, adopting a different way of conducting LDP evaluation that considers alternate models and approaches, in harmony with functionalist tools, is proposed. Next, we briefly introduce qualitative evaluation models and approaches, before presenting our review methods.

Qualitative evaluation models/approaches

Qualitative evaluation approaches have become more prevalent since the 1970s when they were identified as being important for evaluating policy and its purpose (Tayabas, León, and Espino 2014). A qualitative evaluation can show deeper and unexpected

outcomes from interventions and capture what happens during the intervention as well as pre and post (Patton 2015). Whilst the value of such approaches is noted (Lincoln and Guba, 1985; Mertens, 2015), they remain underutilized (Minshew et al. 2021; Spencer et al. 2003). Qualitative evaluation approaches may be particularly suitable in healthcare leadership contexts. For example, qualitative interpretative perspectives on leadership assume that leadership is an emergent process that is experienced in teams, groups, and communities, and that this leadership is socially constructed and distributed in the collective, and not in individuals, as functionalists assume. Leadership is embedded, co-created, and enacted in most healthcare contexts and cultures. Consequently, leadership development is assumed to happen organically as participants learn, interact, build their networks, and develop their expertise in everyday practice and in specific work contexts. Within healthcare settings, leadership can be seen to emerge and develop among the teams of healthcare professionals through informal means, in collaborative project environments. Formal LDPs might facilitate such leadership emergence (Turner 2019). Yet what is developed through LDPs, what happens because of the leadership that is developed, if any, and how do we know that if the development of leadership actually improved both clinical outcomes and satisfactions for patients, providers, and other stakeholders are not fully known, in part due to the limited usage of qualitative evaluation approaches.

Patton (2015) argues that qualitative findings are critical 'to enhance quality, improve programmes, generate deeper insights into the root causes of significant problems, and help prevent problems' (p. 205). This is why qualitative models that help distinguish the specific context and interactions and what goes on in these contexts are central to evaluation practice, as they acknowledge and reveal the relevance of alternative interpretations of a situation to inform change. Such models have the potential to be readily applied in evaluating healthcare LDPs, in conjunction with other functionalist evaluation models and approaches, or as a stand-alone healthcare evaluation practice (Wäscher et al. 2017).

Methods

In this scoping review, two main sources are used to identify the evaluation models and approaches used for evaluating social programmes, policy, and practice, within the programme evaluation literature. First, we searched for the published reviews of evaluation models and approaches within the databases and textbooks, and then we specifically searched for impact evaluations published in the evaluation-focused academic journals, as these sources cover most of the published evaluations that are conducted in various contexts. We draw on both sources, equally, to identify the models and approaches that are the most important and relevant to healthcare organizations. Since scoping reviews are useful when the information on a topic has not been comprehensively reviewed or is complex and diverse (Sucharew and Macaluso 2019), as in the case of LDP evaluation, we sought to hunt for qualitative models and approaches used in a range of different evaluation designs.

First, we sought to identify the published reviews of evaluation models and approaches. Using the databases, Business Source Premier, ABI/INFORM Global, Scopus, and Social Science Citation Index, we deployed the search terms 'leadership development programmes' and 'qualitative evaluation' in combination with the terms

'model', 'approach', "framework', "technique', 'tools', and 'review', to pinpoint potential review papers. With the very few papers that we identified (e.g. Patton 2015; Linzalone and Schiuma 2015; Brandon and Ah Sam, 2014; Contandriopoulos and Brousselle 2012), we recognized the lack of a comprehensive collection and reviews of evaluation models and approaches in the literature. Then, with the support of a subject-specific librarian, and by using the same search terms on Google Books, we identified the textbooks that contain a review of the evaluation models and approaches (marked with an asterisk in references). In the second stage, we examined nine journals that specialize in publishing evaluation studies: *American Journal of Evaluation; New Directions for Evaluation; Canadian Journal of Programme Evaluation; Evaluation; The International Journal of Theory, Research and Practice; Evaluation Review; Evaluation and Programme Planning; The Evaluation Exchange; Practical Assessment, Research & Evaluation; and Journal of Multi-Disciplinary Evaluation*. A total of 22 models were identified. Since our purpose is to help enhance evaluation practice within healthcare organizations, we then purposely select models and approaches using the following criteria: models and approaches are relatively simple, flexible, and in recurrent use, and they can be readily applied by healthcare evaluators, with relative ease and training. This resulted in a total of 17 models for appraisal, as ordered in Tables 3, 4,5. In each table, using Mabey's (2013) framework, the models and approaches that are aligned with interpretivist, dialogic, and critical perspectives are grouped. These models and approaches rely on qualitative data collection methods to guide evaluators on how to go about undertaking an LDP evaluation, what steps must be taken and how to engage with stakeholders. Each table includes the assumptions that underlie each model/approach, and the necessary steps involved in applying them, whilst highlighting their strengths, weaknesses, and context suitability, along with the theorist or proponent of each.

Results

A range of qualitative models and approaches have been identified in the programme evaluation literature (see Tables 3,4,5). Since the evidence for the use of these qualitative evaluation models and approaches within the healthcare literature is limited, we highlight here the potential value of their application when evaluating healthcare LDPs.

Models/approaches aligned with the interpretative perspectives of leadership development evaluation

Mabey (2013) clarifies that leadership from an interpretivist position shifts the emphasis from an individual to a shared approach, assumes that leadership is often culturally situated and highlights that the role of leader may be a fluid one. In the interpretative position, how leaders make sense of their role, the situation and that of others is also significant. The rationalization of events in a post hoc development of learning is captured. For the LDP to mirror this, and the methods used, it may mean that development is taking part in real-time in the workplace as opposed to an external space, and that leadership is emergent, collaborative, and not pre-determined. Exploring the LDP

members' response to the actions that have occurred in a LDP, then their lived experience during and after the event will form a critical part of learning. For the evaluators that are capturing this, it can be challenging.

In the search, six evaluation models in the interpretive space were identified (see Table 2). These are the Culturally Responsive (Frierson, Hood, and Hughes 2002), Culturally Competent (Chouinard and Cousins 2007), Goal-free Evaluation (Scrivens 1999), Connoisseurship Evaluation (Eisner 1997), the Photovoice method (Yuan and Feng 1996) and an open-systems based, EvaluLEAD framework (Grove, Kibel, and Haas 2007).

Key themes and opportunities are the exploration and implications of the culturally grounded nature of leadership development. The opportunity to discover what is going on in a situation and how leadership practice is embedded and institutionalized within a specific healthcare system in a given context is offered (Esmail, Kalra, and Abel 2005; Kalra, Abel, and Esmail 2009). Moving beyond the predetermined goals and objectives in order to engage with other outcomes and their implications can reveal the unexpected consequences of the LDP (Scriven 1999). These models can help capture key stakeholders' views on what constitutes leader excellence (Eisner 1997) and help HRD professionals understand how leadership emerges in teams, groups, and networks (Yuan, and Feng 1996). EvaluLEAD framework advocates the use of evocative forms of inquiry (that employs tools such as stories, journals, visual images, and diaries), along with evidential forms of inquiry (that rely on quantitative data) to capture qualitatively different outcomes, at multiple levels.

Models/approaches aligned with the dialogic perspectives of leadership development

A dialogic perspective on leadership assumes that leadership is a 'discursive accomplishment' that is 'continually in a state of becoming as opposed to anything more fixed or stable' (Mabey 2013, 366), and leadership learners become who they are, on the basis of the stories that they tell of themselves and of their organization, as they engage in everyday conversations. Discursive leadership points to multiple, fragmented, intertextual, and constantly shifting leadership identities that are enacted in specific sociohistorical contexts. Consequently, developing such leadership is assumed to happen as individuals craft their own identities through framing and reframing of personal and organizational stories. Leadership development then becomes fluid, fragmented, and overlapping, and is sometimes contradicting growth of their self in each context.

The review identified seven evaluation methods and approaches that are aligned with the dialogic perspective (see Table 3). These are the Success Case Method (Brinkerhoff 2005), Most Significant Change Method (Dart and Davies 2003), Stakeholder-based evaluation (Mark and Shortland 1985), Collaborative evaluation (Rodriguez-Campos 2012; O'Sullivan 2012), Utilization-focused evaluation (Patton 1997), Illuminative evaluation, (Parlett and Hamilton 2017), and Appreciative Inquiry (Cooperrider and Whitney 2005; Ludema, Cooperrider and Barrett, 2006).

Although these models/approaches were originally conceived and mostly used as tools for improving a programme or for learning from it, they can serve as effective tools to understand the stories that leadership learners tell of themselves. They enable the recognition of other actors that are participating in dialogic discourses on leadership.

Table 2. Methods/approaches aligned with interpretative perspectives on leadership and leadership development.

Evaluation model	Approach	Underlying Assumption(s)	How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/ advocate
1.Culturally responsive evaluation (Frierson, Hood, and Hughes 2002) 2.Culturally competent evaluation (Chouinard and Cousins 2007)	Addresses issues relating to the influence of cultural context of LDPs. Culturally competent evaluators are responsive to the influence of a programme's cultural context when designing, implementing and reporting programme value.	Culturally-defined values and beliefs lie at the heart of this evaluation. Leadership learners' socio-political, demographic, and contextual dimensions, locations, perspectives and characteristics of culture are fundamental to evaluation.	• Develop cultural competence in evaluators so that issues of culture and race are addressed in evaluation. • Educate evaluators in learning about the role Context plays in leadership learning and transfer, so that they engage with stakeholders meaningfully. • Ensure that only those who are knowledgeable about individual, team, organizational, national culture evaluate LDPs, by using culturally appropriate tools.	Suitable for evaluating LDPs that are targeting leader development of specific groups of learners (e.g., from black and minority ethnic communities, or in hospitals where cultural responsiveness is viewed as foundational to more equitable policies and procedures.	• Places relatively greater emphasis on the context, which helps an evaluator to bring out a culturally grounded understanding of a LDP. • Challenges evaluators to reflect on power dynamics and sensitize them to racial and identity issues of learners and their learning.	May be more challenging to develop and retain culturally competent evaluators.	Hood,
Hopson, and Frierson (2005).**TEXT MOVED FROM PREVIOUS COLUMN**							

(Continued)

Table 2. (Continued).

Evaluation model	Approach	Underlying Assumption(s)	How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/ advocate
3.Goal-free evaluation	Evaluating a LDP without knowing the goals/objectives of the programme. The evaluator asks 'what are the actual outcomes of a LDP?' and studies the LDP's observable outcomes in relation to learners' needs.	A goal-oriented search creates an occupational tunnel vision that tends to ignore the emerging, unintended outcomes of a LDP, therefore, conduct evaluation without looking at the intended aims and objectives.	• Gather data on a wide range of *actual effects* or outcomes of a programme and compare them with the *actual needs* of the programme participants. • Do not look in the direction of intended effects/outcomes. • Intentionally avoid reading of programme documents, commissioning briefs, and discussing with programme staff.	All types of healthcare contexts; By combining goal-free and goal-based evaluations, preferably with two different evaluators, inquiries can bring out comprehensive information of programme impact.	Can reveal unanticipated outcomes; Helps eliminate the possible perceptual biases introduced into an evaluation by knowledge of pre-determined objectives. Enables evaluators to avoid an 'occupational tunnel-vision' of looking only in the direction of programme goals. Helps maintain evaluators' independence and objectivity.	Programme staff might 'keep dropping hints about programme goals" (Patton 2015, 206) and therefore, evaluators need to be cautious in maintaining their objectivity. Evaluations can reveal undesired, negative impacts too. Funders and commissioners may want to have some evidence in support of their goals, and producing something contrary to their expectations, may not be encouraged. Evaluators need to defend their analysis.	Scriven (1999)

(*Continued*)

Table 2. (Continued).

Evaluation model	Approach	Underlying Assumption(s)	How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/ advocate
4.Connoisseurship Evaluation	Based on an expert's perceptions and judgements on what constitutes excellence. An expert (i.e. connoisseur), critically appraises a LDP and describe what is going on in a programme (its factual, emotional and aesthetic aspects), interpret the findings in light of theoretical and practical knowledge, and makes evaluative conclusions. • Experts rely on their own expert experience in leader/ship development and draw heavily on their own judgement about what constitutes excellence or leadership in a given context.	When an expert directly observes what is going on in a programme, they can 'see' what is not obvious to lay persons. Like an 'art critic', they can immerse themselves in the programme context and reveal rich descriptions of that programme and its outcomes. They might use methods such as direct observation and immersion in a specific healthcare context.	• Invite, for example, an industry leader, who sets direction, align people and maintain commitment (Drath et al. 2008) in a real-world organization, to look at the inputs, processes, outputs, and outcomes of your LDP. As connoisseurs, they can make evaluative descriptions that are of value to decision-makers. • An experienced evaluator or a leadership scholar can also act as a connoisseur.	Suitable when such experts are available in a programme context (for e.g. in a teaching hospital, where experienced leaders/ leadership scholars are available to act as experts).	• Experts could give us useful impact information, with greater specificity, in a language decision-makers understand. • They can reveal unexpected, unintended, and undesired outcomes too. • They can also do a meta-evaluation (an evaluation of evaluations) and identify the strengths and weaknesses of evaluation practices.	Possible perceptual bias. The credibility of the findings is dependent on the capabilities and expertise of the critical expert, who leads the evaluation.	Eisner 1997
5.Photovoice Method	A participatory evaluation method that uses photos and captions to identify, from learners' perspective, the most significant accomplishments of a LDP.	Any learner, irrespective of their linguistic abilities should participate in programme evaluation.	Ask them to bring an image OR co-create an image/ a collage to illustrate programme impact.	Suitable to identify the development of leadership in networks, communities, and collaborations (e.g. in community care settings).	An effective tool to identify the impact of community-based, leadership development programmes. People with limited experience in evaluation, and patients and carers of various kinds can participate in the process.	Data analysis may be challenging OR may require skilled personnel.	Wang, Yuan, and Feng (1996)

(Continued)

Table 2. (Continued).

Evaluation model	Approach	Underlying Assumption(s)	How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/ advocate
6. EvaluLEAD Framework	Presented as a holistic design that promotes the practice of looking for results in areas leadership learners interact. A sharper focus on what the LDP is doing to trigger a change. Encourages evaluators look for outcomes at individual, organizational, community and societal levels.	LDP participants engage in transfer-actions, in various organizational and community settings in which leadership is enacted, and therefore, outcomes can be identified in multiple levels. By mapping out the types of results, the levels in which the results could emerge, the appropriate data collection methods for each context, more meaningful evaluations can be done.	• Determine purposes, assumptions, and expectations of key stake-holders, evaluation processes based on how leadership is interpreted in a context • Map out all possible types of outcomes a LDP is expected to produce (e.g. episodic, developmental, and transformative changes) • Identify the 'Domains of impact' (i.e. individual, organizational, societal, and community levels) • Employ a range of appro-priate 'evidential' and 'evocative' tools to cap-ture, document, and com-municate the outcomes.	Suitable to evaluate large scale, investment heavy LDPs that expected to produce system, community and societal level outcomes in complex healthcare systems. Encourages the use of 'evocative forms of inquiry' through the use of qualitative surveys, anecdotes, journals and video diaries, in addition to quantitative methods.	Provides a language to talk about various result types, their domains, and the associated methods for outcomes and capture and reporting. It prescribes a systematic way of undertaking an evaluation. As it is based on an open-systems perspective, it recognizes the multidimensionality of relationships between inputs, processes and outcomes (as opposed to linear logical models, promoted in theory-based evaluation approaches).	The framework encourages the use of two different forms of inquiry to identify three result types, in three domains, and thus eighteen activities (2x3x3 = 18) have been identified as prototypical evaluation activities. These activities may be too complex to implement in full, for novice evaluators. Application of this framework might require methodological expertise in data collection and analysis. Published examples of this using this framework are limited.	Grove, Grove, Kibel, and Haas 2007; Black and Earnest 2009.

Table 3. Methods/approaches aligned with dialogic perspectives on leadership and leadership development.

Evaluation model	Approach	Underlying Assumption(s)	• How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/ advocate
7.Success Case Method (SCM)	A process of highlighting the impact variation experienced by successful and unsuccessful adopters of a programme, using story-telling techniques.	Impact must be understood in a context of a performance management system. If there is an impact identified, then a system should have contributed to that impact, in a particular context. Stories are helpful tools to bring out this information.	• Through a survey, identify a small group of success cases (individuals who experience positive effects of aLDP) and interview them to understand what worked for them. • Interview another sample of people who found no or negative value of the programme. Document what did not work for them and why. • Compare the stories and report the results.	Suitable to use even in early stages of a programme. Suitable for Clinical leadership development programmes, (where there is paucity of rigorous evaluations in the UK, Hofmann and Vermunt 2017)	Relatively simple, cheap, quick way of finding out what works and what doesn't, with reasonable rigour and accuracy.	Selection of cases need to be done objectively. Some stories may not be that compelling to tell the full impact of a LDP. There may be a need to combine this method with another method.	Brinkerhoff (2005)
8.Stakeholder-based evaluation (Mark and Shortland 1985) 9. Collaborative evaluation (O'Sullivan 2012; Rodriguez-Campos 2012)	Programme participants become active 'co-evaluators' by allowing them to make reflective conclusions about programme outcome and impact. In most cases, it is an evaluator-led inquiry, in which participants take part in varying degrees, in different evaluative tasks. Participatory evaluators 'jointly share' control of the evaluation.	Stakeholder engagement increases the use of evaluation findings; ensures that findings are rooted in the data; enhances capacity building; and mobilizes social action.	• Create ongoing engagement with stakeholders. • Act as a facilitating collaborator before, during and after the evaluation, so that the entire experience becomes a capacity-building exercise done in trust, and with respect to all those who are involved. • Use simple qualitative methods such as observation and interviews, so that lay people can understand and use them.	A wide range of people, including service users, can be invited to become collaborators in evaluation. Suitable for most healthcare contexts including primary, non-profit hospitals, and community hospitals.	It can bridge diverse perspectives on programme impact; the process of getting involved in various tasks in evaluation, can develop participants' capabilities in undertaking systematic inquiries, which can have broader impact in their own personal and professional lives. This capacity building can also help create self-evaluating, continuously learning organizations.	• Participation must be real authentic and meaningful; if not, the process will be perceived as tokenistic, patronizing, lip-service and the process will fail. • All need to be aware of the power-relations exist between the parties and care needs to be taken to focus on conducting evaluations professionally.	Cousins and Earl 1992

(Continued)

Table 3. (Continued).

Evaluation model	Approach	Underlying Assumption (s)	• How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/ advocate
10.Most Significant Change (MSC) method	A dialogical, story-based evaluation tool. A panel of key stakeholders collectively search for significant change stories and jointly decide the value of collected stories, in a care context.	When change is collectively recognized and acknowledged by key stakeholders, evaluation becomes more meaningful.	• Ask the participants, 'Looking back … (over the last month), what do you think was the most significant change in … (include a domain name, or level of impact). • Collect stories • Let a panel of key stakeholders select the most significant stories • Verify those stories' reliability. • To add legitimacy and rigour, do secondary analysis of contents across the stories, and do 'meta-monitoring' – i.e. monitor how themes change across stories over a period of time.	Suitable for most healthcare contexts.	• Can be an effective ex-post, summative evaluation tool. • Can capture multi-level outcomes (i.e. individual, team, department, organi-zation, and community). • It has the ability to promote extensive dialogue between key stakeholders. • Stories can be seen as less threatening by lay stakeholders.	• Care needs to be taken to monitor whose stories are selected, what hap-pens to those stor-ies, the recommendations produced, and how the learning is utilized.	Dart and Davies (2003)

(Continued)

Table 3. (Continued).

Evaluation model	Approach	Underlying Assumption(s)	• How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/ advocate
11.Utilization-focused evaluation	Emphasizes the intended uses of evaluation results by intended users. It is based on the belief that evaluations should be judged by their utility and actual use. Therefore, the values of primary intended users of evaluation findings must guide the design of an evaluation exercise. It operates from the premise that evaluation use is too important to be merely hoped for or assumed. Use of evaluation findings must be planned for and facilitated.	If evaluation findings are to be used by intended users, 'usefulness' must guide the entire process of planning, implementation and reporting of an evaluation. Situational Responsiveness of evaluators is the key. Evaluations must be judged by its use. The intended users are more likely to use evaluations if they have been actively involved in evaluation design and implementation.	• Identify a group of primary intended users of evaluation. They should represent the actual users of evaluation. • Involve them in the designing and implementation of evaluation so that they own the outcomes. (i.e., ask them what kind of evaluation they need and select content, methods, theory, and uses that are relevant to them). • Use your personal factor to enhance the use of evaluation findings.	Suitable for LDPs designed for improving a programme (rather than proving value of a programme).	Emphasizes the utility of evaluation findings, through participation of primary intended users. No adherence to any model, procedures, values or standardized recipe-type approaches. Uses eclectic approach to enhance utilization of evaluation.	Turnover of users is an acknowledged problem in the literature. Replacement of users delay the process. This approach is vulnerable to bias and corruption by the users' group. Even when evaluations are designed and implemented through consultations and collaborations, it is still possible that the findings are not 'used' in certain contexts, due to power differences and internal politics.	Patton (2008)

(Continued)

Table 3. (Continued).

Evaluation model	Approach	Underlying Assumption(s)	• How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/advocate
12.Illuminative Evaluation	A flexible evaluation relying on naturalistic method and mixed method approaches; 'Going beyond programme objectives, it seeks to illuminate essential programme issues and the complex contextual information of a LDP. This approach 'seeks to address and illuminate a complex array of questions.' (Parlett and Hamilton, 1976, p.144).	• An anti-positivist in orientation. • Thoroughly context-bound • Multi-faceted in perspectives • Illustrative of the mismatch that often occurs between rhetoric and action • Concerned with revealing recognized and recognizable reality (Burden 2017, 298)	Consider: **SPARE** Investigate the **S**etting with a view to construct a rich picture of programme setting; **P**lan: what is the programme, whose idea is it, how the intended outcomes are conceptualized to come about. **A**ctions: What actions have been taken, is there a mismatch between what is planned and what is enacted **R**eactions: what are the reactions of learners about the current state of the programme. **E**valuation: the illuminative evaluator is in constant reflection on what is going on and in regular dialogue with key stakeholders. (Burden 2017, 299)	All 'place-based' LDPs, sponsored in for example, community hospitals, federal government hospitals and non-federal psychiatric/ long-term care hospitals.	When there is a need to understand complexities of an on-going programme, this approach can be used to illuminate the complex realities surrounding that programme.	Published examples of scholars using this approach for evaluating LDPs are relatively rare, although its uses in other healthcare contexts exist (see McFarlane, 2004).	Parlett and Hamilton ()

(Continued)

Table 3. (Continued).

Evaluation model	Approach	Underlying Assumption(s)	• How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/advocate
13. Appreciative inquiry	Seeks to understand what is best in a programme, to create a better system. This approach goes beyond the pre-determined programme objectives. It involves LDP participants asking each other about the times they felt most alive, most excited and hopeful in a programme context. Programme participants generate stories and metaphors about the positive aspects of their work-life; they share their hopes and dreams for making their organizations flourish.	Help participants focus on the possibilities and opportunities for the future, by asking them certain types of questions; help them focus on the positives. The focus becomes reality because programme participants can see what is good and possible in the future.	• Discover: Ask participants about a time when they felt most alive and fulfilled or most excited about their learning and development. Aim to generate examples, stories and metaphors about the positive aspects of a programme. • Dream: Help them articulate what else is possible when applying the learning gained; what impacts it might create for their teams, organizations, and stakeholders? • Design: Ask participants to articulate what needs to be started/stopped/changed as a result of a programme? What can they do collectively as leadership-learners, to make their dreams come true in their work context? • Destiny – Ask them to commit to delivery results, continuous learning, and innovation in leadership practice. • Document and report learning gains.	All types of programme contexts, where collective leadership development is aimed at system levels (e.g. evaluating a structured Nursing leadership development programme for succession planning, Ramseur et al. 2018).	Powerful questioning strategies, prescribed in this approach can be incorporated well in evaluation. Instead of focusing on the negatives, with a pessimistic critical lens, this approach focuses on participants' core capacities, strengths and successes, relying on their imagination and creativity. Through insightful questioning, it helps participants' core strengths, successes and their beliefs on what is possible. This process of dialogue can happen between an evaluator and programme participants; it can be also used when participants themselves interview each other.	Evaluators may uncritically, over-emphasize the positives, while ignoring what did not work.	Cooperrider and Whitney 2005

Table 4. Methods/approaches aligned with critical perspectives on leadership and leadership development.

Evaluation model	Approach	Underlying Assumption(s)	How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/ advocate
14.Empowerment Evaluation (Fetterman, Kaftarian, and Wandersman A 1996) 15. Transformative evaluation (Mertens 2009)	Fosters 'self-determination' among those who participate in an evaluation process. In some cases, developing them to conduct their own evaluations and become self-sufficient. In these, programme staff, participants and beneficiaries of the programme control evaluation.	Capacity-building, self-determination and empowerment mobilize social action.	• Form 'empow-erment-partnerships' with those who collaborate in the evaluation. • Serve as critical friends and coa-ches to help keep evaluation processes rele-vant, rigorous, responsive, and on budget. • Participants make key deci-sions to make evaluations aligned with their contexts, and goals and purposes of evaluation.	Suitable in contexts where patients voice is at the core of decision making.	Stakeholders are not considered as 'subjects' of evaluation. It focuses on participant empowerment through evaluation process. Transformative evaluation provides guidance in designing and implementing evaluation that promote social justice aims.	Very useful in a resource-constrained healthcare organizations. A wide range of published works give case examples of empowering stakeholders to do their own evaluations, in various contexts.	Fetterman 2013

(Continued)

Table 4. (Continued).

Evaluation model	Approach	Underlying Assumption(s)	How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/ advocate
16.Feminist Evaluation	*Gendered evaluation* is contingent on the meanings given to the terms *sex* and *gender*. In feminist evaluation, the evaluator is mindful of gender issues, the needs of women and the promotion of establishing equal rights and opportunities.	• Gender inequities are one manifestation of social injustice. • Sex and gender are inextricably linked to race, class, ethnicity, age, ability and sexual identity. By understanding the dynamics at play at the nexus of these identities from a feminist perspective, can lead to more powerful ways of doing research and evaluation	• Acknowledge and examine the structural nature of inequities, beginning with gender as your starting point. • Be aware that an evaluators' personal experiences, characteristics, and perspectives shape their political stance; • Acknowledge discrimination based on gender is systemic and structural. • The purpose of knowledge is action and therefore, action and advocacy are morally and ethically appropriate responses of an evaluator.	Suitable for contexts that are attentive and responsive to gender and issues related to gender equality and women. Suitable for school-based healthcare programmes (e.g., Seigart 2014).	It gives voice to underrepresented voices and propels action to work for social justice. These social-justice focused models raise questions about the roles and responsibilities of an evaluator and examine the	There may be a need to clarify the distinctive contribution this approach makes to evaluation in a given context and how this approach contributes to social justice aims. A rigid approach may lead an evaluator to overlook or neglect potentially, even more significant outcomes.	Brisolara, Seigart, and SenGupta (2014)

(Continued)

Table 4. (Continued).

Evaluation model	Approach	Underlying Assumption(s)	How to apply?	Suitability for healthcare contexts	Strengths /benefits	Possible Limitations	Prominent theorist/ advocate
17.Horizontal Evaluation	Combines self-assessment by local participants and external review by peers.	• The involvement of peers neutralizes the power imbalances that prevail in traditional external evaluations, creating space for new learnings.	• Identify that which is to be evaluated. • Recruit and ensure participation of an appropriate group of local participants and external reviewers. • Design a workshop, find a facilitator and collectively develop evaluation criteria. • Arrange field visits so that all can see how leadership happens in a care context. • For each evaluation criterion, the two groups separately identify and present strengths, weaknesses, and suggestions for improvement. • Write up report and use findings.	Suitable for all leadership contexts with power differences.	• Overcomes external evaluators that limit new learnings. • Facilitates the building of trust, social learning, and a sense of community • Promotes ownership of results • Help local participants accept critical feedback and observations	Mainly used as for formative evaluation (i. e. to improve a situation, not to prove something).	Thiele et al. 2007

Models/approaches aligned with the critical perspectives of leadership development

A critical perspective on leadership sees LDPs as the means of promoting a blind acceptance of norms and the status quo, and of propagating the knowledge and ideas that serve the interests of a powerful elite while treating participants as passive consumers of what is being done to them in a formal classroom, mostly by those who represent a dominant group in a context, and in some contexts by male presenters (Sinclair 2009; Ford 2015). In some cases, LDPs, as 'a covert means of perpetuating political elite domination' (Tomlinson, O' Reilley & Wallace, 2013, 81), could even become exploitative by masking the power relations that exist in care settings (Currie and Spyridonidis, 2016; 2019), and by deflecting participants' attention from emotional, structural, and political barriers to systemic change, whilst focusing exclusively on heroic, personal transformation, and individual successes as markers of leader development (Willmott 1997; Alvesson and Spicer 2012)

The review provided four evaluation approaches that are aligned with the critical perspective of leadership (Table 4). These are the Empowerment Evaluation (Fetterman, Kaftarian, and Wandersman A 1996), Transformative evaluation (Mertens 2009), Feminist evaluation (Bustelo 2017; Brisolara, Seigart, and SenGupta 2014), and Horizontal Evaluation (Thiele et al. 2007). These approaches help us understand 'the dialectical asymmetries, situated interrelations and intersecting practices of leaders and followers' (Collinson 2017, 272) in care contexts characterized by politics and power differences.

The key themes are power dynamics, and the nature and consequences of gender inequality, whilst noting the racial, gender, and identity issues of learners and their learning (Smith and Gosling 2017; Neubert and Palmer 2004). These models help evaluators recognize the complexity of leadership learning and development and acknowledge the power differences held by various groups including healthcare commissioners, senior management, and marginalized groups who may be involved in the formal and informal LDPs. Offering HRD practitioners further knowledge of how resistance and cynicism flow among different groups in hospital settings, and how leadership is seen as an emerging outcome of inter-dependant relationships and networks within a political context. By focusing on the dialectics of control and resistance and the ideological aspect of leadership, these approaches enable evaluators and HRD practitioners to ensure that the unique characteristics of leadership-learners are not lost in the development process, and to consider if LDPs promote the establishment of equal rights and opportunities.

Discussion

In this paper, we argue that rather than reducing the practice of impact evaluation to the application of just one or two popular evaluation models, healthcare LDP evaluators could recognize the diversity of qualitative models and approaches that are available in other fields of practice, including that of programme evaluation. The seventeen programme evaluation models and approaches reviewed in this paper have the potential to answer a wide range of impact questions. They can assist HRD practitioners to discern

how these models and approaches can be used exclusively on their own, or in conjunction with other approaches for mutual facilitation and complementarity in specific contexts. Interpretive culturally responsive (CRE) and culturally competent evaluation (CCE) models can provide HRD practitioners with data that moves beyond narrow assumptions about leadership and leadership development. For example, as in the case of Kalra, Abel, and Esmail (2009), a healthcare organization may endorse and champion leaders from Black, Asian, Minority Ethnic (BAME) backgrounds and support their attendance on an LDP. A functionalist evaluation model might show evidence of programme satisfaction. However, other workforce data show that these BAME leaders experience continued under-representation in hospital boards and racial inequality at work (O'Dwyer-Cunliffe and Russell, 2020), and they tend to leave their posts and the organization sooner than their white peers. To discover why this is the case may be significant, and culturally responsive evaluations could help us understand how the programme influences next steps for LDP attendees (if at all) and if there is any useful learning about the programme and the organizational context, which can help address the issues faced by BAME leaders. Such methods enable different perceptions, views, and values to be made accessible for discussion, identifying the individual needs of the stakeholders (Wäscher et al. 2017) to create a shared perspective.

Dialogic models such as those used in collaborative evaluation can be important when attempting to explore leadership in teams – a key aspect of integrated healthcare but in reality, often under explored or examined in the LDP process for impact or outcomes (Pallesen et al. 2020). Similarly, critical approaches, such as the empowerment evaluations have been successfully used in several healthcare quality improvement contexts. For example, to evaluate the effectiveness of HIV prevention programmes (Phillips et al. 2019), to reduce hospital admissions through improved diabetes care (Wandersman 2015), and to facilitate power-sharing and joint decision-making among nurses and families (Strober 2005), empowerment evaluations have been used effectively. In all these studies, the chosen critical evaluation approach has provided a better understanding of work processes and institutional arrangements in healthcare settings, and helped evaluators build capacity at these organizations to foster a learning-focused community.

The qualitative models and approaches reappraised here may be helpful in advancing LDP evaluation practice in healthcare contexts. Equally, the findings from such evaluations might help HRD practitioners to gain added value and make better sense of leadership learning situations, along with learners' differences in beliefs, intentions, and values, as well as the social, cultural, and emotional factors that affect leaders and leadership development.

Implications for HRD research

Recently, HRD scholars argued for integrating temporal dimension in LDP evaluations (Joseph-Richard, Edwards, and Hazlett 2021) and highlighted the advantage of such integration for designing, implementing, and using evaluations. By knowing the timing and duration of outcomes, HRD practitioners can make a more realistic estimation of the scope of personal and relational changes that could be observed in healthcare contexts and create more-efficient learning and development investment strategies. Unfortunately, the models and approaches reviewed here offer limited guidance on

how temporality could be integrated in evaluation designs. More theorization, research, and explicit guidance are needed to help HRD practitioners in this area. Published examples of applying these models/approaches, which present a vivid picture of programme contexts, the evaluation questions used, the appropriateness of the methods employed, the processes data analysis, challenges faced, and lessons learned by programme staff could enhance our understanding of how leadership and leaders are developed, and outcomes are experienced in multi-cultural, cross-border healthcare contexts. Since empirical evidence that supports the effectiveness of these models and approaches, when applied in healthcare contexts, is very limited, meta-evaluation works that investigate the effectiveness of these tools might reveal what works, for whom, when, where, why, and how. Finally, since the under-use of evaluation findings has been well recognized (Long 2006), we also need to show how evaluation findings can be utilized in healthcare contexts so that investments in LDP evaluations could be justified in terms of the learning-gains acquired at personal, professional, and organizational levels.

Implications for HRD practice

Given the wide range of models and approaches that are underutilized in the current practice, it may be useful to highlight one or two models as better tools for the job. However, considering the rich variation in changing healthcare contexts, it becomes challenging both to single out individual models/approaches as more effective tools than others, and to promote them as more suitable to certain contexts. Every healthcare context is unique. Such recommendations may even be considered as overly prescriptive. We endorse what Inouye, Yu, and Adefuin (2005) emphasized in their evaluation commissioners' guide that evaluators must 'take into account potential cultural and linguistic barriers', re-examine 'established evaluation measures for cultural appropriateness', and/or incorporate 'creative strategies for ensuring culturally competent analysis and creative dissemination of findings to diverse audiences' (p. 6), particularly when selecting methodological designs and tools. In line with Patton (2008), we believe that a useful, practical, ethical, and accurate evaluation 'emerges from the special characteristics and conditions of a particular situation – a mixture of people, politics, history, context, resources, constraints, values, needs, interests, and chance' (p. 199). However, we can point out that combining certain complementing models and approaches may be fruitful in certain contexts.

Although the four categories are presented as distinctly different sets of evaluation models and approaches, they are by no means pure types. These categories are most distinct with respect to the key assumptions on how each of them should be implemented in a given context. Significant differences also exist with respect to the locus of evaluators' power. However, there are many aspects of these approaches that are quite similar, depending on which category of approach one adopts. For example, although stakeholder-based evaluation is the approach most visibly concerned with stakeholder engagement and participation, there is a value position that is at least implicit in most of the approaches. Empowerment evaluation approaches are quite explicit about the centrality of power relations, and it is important to recognize who conducts an evaluation, when, where, how long, and why, even when using other approaches that are not explicitly emphasizing the role of power.

Mixing these approaches, in ways that are suitable for the given healthcare context is the key. For example, at the start of an evaluation project in a small tertiary referral hospital, the goal-free evaluation approach (interpretative) could be used for exploratory purposes (as Scriven 1999 himself proposed), followed by the use of the Success Case Method (dialogic) for collating evidence for impact and programme improvement. In a community (geriatric) hospital, mixing a culturally responsive evaluation (interpretative) with a stakeholder-based evaluation (dialogic) may be useful. Such purposeful mixing of models and approaches in evaluation designs could help HRD professionals draw on the potential strengths of each method while simultaneously mitigating their weaknesses. However, such decisions need to be taken only in full appreciation of the increasingly diverse, complex, and adaptive healthcare systems, which generally vary across international borders (Plsek and Wilson 2001; Greenhalgh and Papoutsi 2018). We recommend that HRD practitioners adopt an eclectic approach that is characterized by an awareness of the full range of options that are available in the literature, and to be guided by a commitment to methodological appropriateness to a given situation, when designing and implementing evaluations (Patterson et al. 2017).

Limitations

This review is limited in that it focuses only on qualitative methods and approaches, found primarily in programme evaluation literature, and future research could find newer tools that are suitable for healthcare contexts in other specialisms, including behavioural economics, neuroscience, and organizational psychology. In selecting these models and approaches, a practice orientation, applicability to healthcare contexts, and the word limits guided our decision. As a result, a few other taxonomy-based models that essentially extend Kirkpatrick's work (e.g. Hamblin 1974; Kaufman and Keller 1994) and approaches, such as responsive evaluation (Stake 1975), responsive constructivist evaluation (also known as fourth-generation evaluation, Guba and Lincoln 1989), democratic evaluation (MacDonald 1993; Picciotto 2015), and developmental evaluation (Patton 1994), among others, have not been included. These approaches also rely heavily on qualitative methods and can be used to evaluate LDPs, although published examples that use these approaches are rare. We believe that practitioner-focused descriptive studies on what actually happens when these models and approaches are utilized, either individually or in combination with other tools in health settings, would be beneficial.

Conclusions

Evaluating leadership development in healthcare contexts is difficult and complex. Rich theorizing and generative learning in the field of LDP evaluations are slowly increasing. Paying more attention to, and pragmatic adaptation of, alternate models and approaches reviewed in this paper (as opposed to relying on a few popular programme evaluation models and approaches that are based on functionalist assumptions), HRD scholars and practitioners could demonstrate the value, if any, of LDPs in healthcare contexts. The sixteen evaluation models/approaches by themselves, we believe, provide a set of rich additional tools for HRD practitioners, working in healthcare contexts across the world. To the extent that these tools are applied, evaluations are published and healthcare

leadership outcomes are convincingly demonstrated in terms of patient outcomes, continuing our evaluation efforts is essential and certainly to be encouraged. We acknowledge that evidence for integrating questions about the timings, duration, and the speed of leadership development outcomes in LDP evaluations is seldom found in these models and approaches. However, applying interpretative, dialogic, and critical models and approaches, either on their own or in combination with suitable tools, can not only provide the much-needed *direction* for designing evaluations but also they can convey powerfully to programme stakeholders the richness of programme impact in ways that are essentially experiential, contextual, participatory, and collaborative.

Disclosure statement

Funding

ORCID

Dr Paul Joseph-Richard ⓘD http://orcid.org/0000-0003-4176-419X
Professor Janet McCray ⓘD http://orcid.org/0000-0001-9159-8770

References

Alliger, G. M., S. Tannenbaum, J. W. Bennett, H. Traver, and A. Shortland. 1997. "A Meta-analysis of the Relations among Training Criteria." *Personnel Psychology* 50 (2): 341–358. doi:10.1111/j.1744-6570.1997.tb00911.x.

Alvesson, M., and A. Spicer. 2012. "Critical Leadership Studies: The Case for Critical Performativity." *Human Relations* 65 (3): 367–390. doi:10.1177/0018726711430555.

Anderson, L. 2010. "Talking the Talk–a Discursive Approach to Evaluating Management Development." *Human Resource Development International* 13 (3): 285–298. doi:10.1080/13678868.2010.483817.

Andersson, T. 2015. "The Medical Leadership Challenge in Healthcare Is an Identity Challenge." *Leadership in Health Services* 28 (2): 83–99. doi:10.1108/LHS-04-2014-0032.

Arthur, W., S. T. Bell, W. Bennett, and P. S. Edens. 2003. "Effectiveness of Training in Organizations: A Meta-analysis of Design and Evaluation Features." *Journal of Applied Psychology* 88 (2): 234–245. doi:10.1037/0021-9010.88.2.234.

Aveling, E.-L., M. Parker, and M. Dixon-Woods. 2016. "What Is the Role of Individual Accountability in Patient Safety? A Multi-site Ethnographic Study." *Sociology of Health & Illness* 38 (2): 216–232. doi:10.1111/1467-9566.12370.

Avolio, B. J., J. B. Avey, and D. Quisenberry. 2010. "Estimating Return on Leadership Development Investment." *The Leadership Quarterly* 21 (4): 633–644. doi:10.1016/j.leaqua.2010.06.006.

Avolio, B. J., R. J. Reichard, S. T. Hannah, F. O. Walumbwa, and A. Chan. 2009. "A Meta-analytic Review of Leadership Impact Research: Experimental and Quasi-experimental Studies." *The Leadership Quarterly* 20 (5): 764–784. doi:10.1016/j.leaqua.2009.06.006.

Bate, P., G. Robert, N. Fulop, J. Vretveit, and M. Dixon-Woods. 2014. "*Perspectives on Context.*" London: Health Foundation London.

Bates, R. 2004. "A Critical Analysis of Evaluation Practice: The Kirkpatrick Model and the Principle of Beneficence." *Evaluation and Program Planning* 27 (3): 341–347. doi:10.1016/j. evalprogplan.2004.04.011.

Bennett, J. 2003. *Evaluation Methods in Research*. London: Continuum.

Birckmayer, J. D., and C. H. Weiss. 2000. "Theory-based Evaluation in Practice." *Evaluation Review* 24 (4): 407–431. doi:10.1177/0193841X0002400404.

Black, A. M., and G. W. Earnest. 2009. "Measuring the Outcomes of Leadership Development Programs." *Journal of Leadership & Organizational Studies* 16 (2): 184–196. doi:10.1177/1548051809339193.

Blume, B. D., J. K. Ford, T. T. Baldwin, and J. L. Huang. 2010. "Transfer of Training: A Meta-analytic Review." *Journal of Management* 36 (4): 1065–1110. doi:10.1177/0149206309352880.

Brandon, P. R., and A. L. Ah Sam. 2014. "Program Evaluation." In *Oxford Handbook of Qualitative Research*, edited by P. Leavy, 471–497. Oxford: Oxford University Press.

Brinkerhoff, R. O. 2005. "The Success Case Method: A Strategic Evaluation Approach to Increasing the Value and Effect of Training." *Advances in Developing Human Resources* 7 (1): 86–101. doi:10.1177/1523422304272172.

Brisolara, S., D. Seigart, and S. SenGupta. 2014. *"Feminist Evaluation and Research: Theory and Practice"*. 1st ed. London: Guilford Publications.

Burden, B. 2017. "Illuminative Evaluation." In *Frameworks for Practice in Educational Psychology: A Textbook for Trainees and Practitioners*, edited by B. Kelly, L. M. Woolfson, and J. Boyle, 291–308. 2nd ed. London: Jessica Kingsley.

Burgoyne, J. G., W. Hirsh, S. Williams, and G. Britain. 2004. "*The Development of Management and Leadership Capability and Its Contribution to Performance: The Evidence, the Prospects and the Research Need."* Nottingham, UK: University of Lancaster: DfES Publications.

Bustelo, M. 2017. "Evaluation from a Gender Perspective as a Key Element for (Re) Gendering the Policymaking Process." *Journal of Women, Politics & Policy* 38 (1): 84–101. doi:10.1080/1554477X.2016.1198211.

Carroll, B. 2019. "Leadership Learning and Development." In *Leadership: Contemporary Critical Perspectives*, edited by B. Carroll, J. Ford, and S. Taylor, 117–137. Los Angeles: Sage.

Chouinard, J. A., and J. B. Cousins. 2007. "Culturally Competent Evaluation for Aboriginal Communities: A Review of the Empirical Literature." *Journal of Multidisciplinary Evaluation* 4 (8): 40–57.

Collins, J., and D. Denyer. 2008. "Leadership Learning and Development: A Framework for Evaluation." In *Leadership Learning: Knowledge into Action*, edited by K. T. James and J, Collins.161–177. Hampshire: Palgrave McMillan.

Collinson, D. 2017. "Critical Leadership Studies: A Response to Learmonth and Morrell." *Leadership* 13 (3): 272–284. doi:10.1177/1742715017694559.

Contandriopoulos, D., and A. Brousselle. 2012. "Evaluation Models and Evaluation Use." *Evaluation* 18 (1): 61–77. doi:10.1177/1356389011430371.

Cooperrider, D. L., and D. K. Whitney. 2005. *"Appreciative Inquiry"*. 1st ed. San Francisco, CA: Berrett-Koehler.

Cousins, J. B., and L. M. Earl. 1992. "The Case for Participatory Evaluation." *Educational Evaluation and Policy Analysis* 14 (4): 397–418. doi:10.3102/01623737014004397.

Currie, G., and D. Spyridonidis. 2016. "Interpretation of Multiple Institutional Logics on the Ground: Actors' Position, Their Agency and Situational Constraints in Professionalized Contexts." *Organization Studies* 37 (1): 77–97. doi:10.1177/0170840615604503.

Currie, G., and D. Spyridonidis. 2019. "Sharing Leadership for Diffusion of Innovation in Professionalized Settings." *Human Relations* 72 (7): 1209–1233. doi:10.1177/0018726718796175.

Dart, J., and R. Davies. 2003. "A Dialogical, Story-based Evaluation Tool: The Most Significant Change Technique." *American Journal of Evaluation* 24 (2): 137–155. doi:10.1177/109821400302400202.

Day, D. V. 2011. "Leadership Development." In *The Sage Handbook of Leadership*, edited by A. Bryman, D. Collinson, K. Grint, M. Uhl-Bien, and B. Jackson, 37–50. London: Sage Publications.

De Déa Roglio, K., and L. Light L. 2009. "Executive MBA Programs: The Development of the Reflective Executive." *Academy of Management Learning & Education* 8 (2): 156–173. doi:10.5465/amle.2009.41788840.

Drath, W. H., C. D. McCauley, C. J. Palus, E. Van Velsor, P. M. G. O'Connor, and J. B. McGuire. 2008. "Direction, Alignment, Commitment: Toward a More Integrative Ontology of Leadership." *The Leadership Quarterly* 19 (6): 635–653. doi:10.1016/j.leaqua.2008.09.003.

Easterby-Smith, M. 1994. *Evaluating Management Development, Training, and Education*. 2nd ed. Hampshire, UK: Gower.

Edmonstone, J. 2011. "Developing Leaders and Leadership in Health Care: A Case for Rebalancing?" *Leadership in Health Services* 24 (1): 8–18. doi:10.1108/17511871111102490.

Edmonstone, J. 2013a. "Healthcare Leadership: Learning from Evaluation." *Leadership in Health Services* 26 (2): 148–158. doi:10.1108/17511871311319731.

Edmonstone, J. 2013b. "What Is Wrong with NHS Leadership Development?" *British Journal of Healthcare Management* 19 (11): 531–538. doi:10.12968/bjhc.2013.19.11.531.

Edwards, G., and S. Turnbull. 2013a. "A Cultural Approach to Evaluating Leadership Development." *Advances in Developing Human Resources* 15 (1): 46–60. doi:10.1177/1523422312467144.

Edwards, G., and S. Turnbull. 2013b. "Special Issue on New Paradigms in Evaluating Leadership Development." *Advances in Developing Human Resources* 15 (1): 3–9. doi:10.1177/1523422312467147.

Eisner, E. W. 1997. "*The Enlightened Eye: Qualitative Inquiry and the Enhancement of Educational Practice.*" New York: Allyn & Bacon.

Ely, K., L. A. Boyce, J. K. Nelson, S. J. Zaccaro, G. Hernez-Broome, and W. Whyman. 2010. "Evaluating Leadership Coaching: A Review and Integrated Framework." *The Leadership Quarterly* 21 (4): 585–599. doi:10.1016/j.leaqua.2010.06.003.

Esmail, A., V. Kalra, and P. Abel. 2005. "*A Critical Review of Leadership Interventions Aimed at People from Black and Minority Ethnic Groups.*" London: Health Foundation.

Fetterman, D. M. 2013. "Empowerment Evaluation: Learning to Think like an Evaluator." In *Evaluation Roots: A Wider Perspective of Theorists' Views and Influences*, edited by M. Alkin, 304–322. London: Sage.

Fetterman, D. M., S. J. Kaftarian, and A. Wandersman A. 1996. "*Empowerment Evaluation: Knowledge and Tools for Self-assessment & Accountability*". Thousand Oaks, CA: Sage Publications.

Finney, L., and C. Jefkins. 2009. "*Best Practice in OD Evaluation.*" West Sussex, UK: Roffey Park Institute.

Fitz-Gibbon, C. T., and L. L. Morris. 1987. *How to Analyze Data*. 1st ed. London: Sage.

Ford, J. 2015. "Going beyond the Hero in Leadership Development: The Place of Healthcare Context, Complexity and Relationships: Comment On" Leadership and Leadership Development in Healthcare Settings–a Simplistic Solution to Complex Problems?." *International Journal of Health Policy and Management* 4 (4): 261–263. doi:10.15171/ijhpm.2015.43.

Ford, K. J., and R. Sinha. 2008. "Advances in Training Evaluation Research." In *The Oxford Handbook of Personnel Psychology*, edited by S. Cartwright and C. L. Cooper, 291–316. Oxford: Oxford University Press.

Frierson, H., S. Hood, and G. Hughes. 2002. "Strategies that Address Culturally Responsive Evaluation." In *The 2002 User-friendly Handbook for Project Evaluation*, edited by J. Frechtling, 63–73. Arlington, VA: National Science Foundation.

Fund, K. 2021. *Integrated Care Systems Explained: Making Sense of Systems, Places and Neighbourhoods*. London: King's Fund.

Funnell, S. C., and P. J. Rogers. 2011. *Purposeful Program Theory: Effective Use of Logic Models and Theories of Change*. San Francisco, CA: Joseey-Bass.

Gagnon, S., and D. Collinson. 2014. "Rethinking Global Leadership Development Programmes: The Interrelated Significance of Power, Context and Identity." *Organization Studies* 35 (5): 645–670. doi:10.1177/0170840613509917.

Greenhalgh, T., and C. Papoutsi. 2018. "Studying Complexity in Health Services Research: Desperately Seeking an Overdue Paradigm Shift." *BMG Medicine* 16 (1): 95. doi:10.1186/s12916-018-1089-4.

Grove, J., B. Kibel, and T. Haas. 2007. "EvaluLEAD: An Open-systems Perspective on Evaluating Leadership Development." In *The Handbook of Leadership Development Evaluation*, 71–110. Francisco, CA: John Wiley & Sons, San.

Guba, E. G., and Y. S. Lincoln. 1989. *Fourth Generation Evaluation*. Newbury Park, CA: Sage.

Ham, C., D. Berwick, and J. Dixon. 2016. *Improving Quality in the English NHS*. London: King's Fund.

Hamblin, A. C. 1974. *Evaluation and Control of Training*. London: McGraw-Hill.

Hartley, J., J. Martin, and J. Benington. 2008. "*Leadership in Health Care. A Review of the Literature for Health Care Professionals, Managers and Researchers.*" National Institute of Health Research: HS&DR. Coventry, UK: University of Warwick. http://www.nets.nihr.ac.uk/projects/hsdr/081601148: .

Hillen, H., H. Pfaff, and A. Hammer. 2017. "The Association between Transformational Leadership in German Hospitals and the Frequency of Events Reported as Perceived by Medical Directors." *Journal of Risk Research* 20 (4): 499–515. doi:10.1080/13669877.2015.1074935.

Ho, M. (2016). "*Investment in Learning Increases for Fourth Straight Year.*" *Retrieved from* https://www.td.org/Publications/Magazines/TD/TDArchive/2016/11/Investment-in-Learning-Increases-for-Fourth-Straight-Year

Hofmann, R., and J. D. Vermunt. 2017. "Professional Development in Clinical Leadership: Evaluation of the Chief Residents Clinical Leadership and Management Programme." *Faculty of Education, University of Cambridge, Working Papers Series* 5: 23.

Holton, E. F. 1996. "The Flawed Four-level Evaluation Model." *Human Resource Development Quarterly* 7 (1): 5–21. doi:10.1002/hrdq.3920070103.

Hood, S., R. Hopson, and H. Frierson. 2005. "*The Role of Culture and Cultural Context in Evaluation: A Mandate for Inclusion, the Discovery of Truth and Understanding*". Greenwich, CT: Information Age.

Hoole, E. R., and J. W. Martineau. 2014. "Evaluation Methods." In *The Oxford Handbook of Leadership and Organizations*, edited by D. V. Day, 167–196. Oxford: Oxford university press.

Inouye, T. E., H. C. Yu, and J. A. Adefuin. 2005. "Commissioning Multicultural Evaluation: A Foundation Resource Guide." In *partnership with Social Policy Research Associates*. Oakland, CA.: California Endowment's Diversity in Health Education Project. https://www.informalscience.org/commissioning-multicultural-evaluation-foundation-research-guide

Jeyaraman, M. M., S. M. Z. Qadar, A. Wierzbowski, F. Farshidfar, J. Lys, G. Dickson, K. Grimes, et al. 2018. "Return on Investment in Healthcare Leadership Development Programs." *Leadership in Health Services* 31 (1): 77–97. doi:10.1108/LHS-02-2017-0005.

Joseph-Richard, P., G. Edwards, and S. A. Hazlett. 2021. "Leadership Development Outcomes Research and the Need for a Time-sensitive Approach." *Human Resource Development International* 24 (2): 173–199. doi:10.1080/13678868.2020.1815155.

Kalra, V. S., P. Abel, and A. Esmail. 2009. "Developing Leadership Interventions for Black and Minority Ethnic Staff: A Case Study of the National Health Service (NHS) in the UK." *Journal of Health Organization and Management* 23 (1): 103–118. doi:10.1108/14777260910942588.

Kaufman, R., and J. M. Keller. 1994. "Levels of Evaluation: Beyond Kirkpatrick." *Human Resource Development Quarterly* 5 (4): 371–380. doi:10.1002/hrdq.3920050408.

Kennedy, F., B. Carroll, and J. Francoeur. 2013. "Mindset Not Skill Set Evaluating in New Paradigms of Leadership Development." *Advances in Developing Human Resources* 15 (1): 10–26. doi:10.1177/1523422312466835.

King, E., and P. Nesbit. 2015. "Collusion with Denial: Leadership Development and Its Evaluation." *Journal of Management Development* 34 (2): 134–152. doi:10.1108/JMD-02-2013-0023.

Kirkpatrick, D. 1996. "Great Ideas Revisited." *Training & Development* 50 (1): 54–60.

Kuluski, K. R., Reid, R. J. R.j, and G. R. Baker. 2020. "Applying the Principles of Adaptive Leadership to Person Centred Care for People with Complex Care Needs: Considerations for Care Providers, Patients." *Caregivers and Organizations Health Expectations* 24 (2): 175–181. doi:10.1111/hex.13174.

Lacerenza, C. N., D. L. Reyes, S. L. Marlow, D. L. Joseph, and E. Salas. 2017. "Leadership Training Design, Delivery, and Implementation: A Meta-analysis." *Journal of Applied Psychology* 102 (12): 1686–1718. doi:10.1037/apl0000241.

Leach, L., B. Hastings, G. Schwarz, B. Watson, D. Bouckenooghe, L. Seoane, and D. Hewett. 2021. "Distributed Leadership in Healthcare: Leadership Dyads and the Promise of Improved Hospital Outcomes." *Leadership in Health Services* 34 (4): 353–374. doi:10.1108/LHS-03-2021-0011.

Lincoln, Y. S., and E. G. Guba. 1985. "*Naturalistic Inquiry.*" Beverly Hills, CA: Sage Publications.

Linzalone, R., G. Schiuma, and P. Harri Laihonen. 2015. "A Review of Program and Project Evaluation Models." *Measuring Business Excellence* 19 (3): 90–99. doi:10.1108/MBE-04-2015-0024.

Long, A. 2006. "Evaluation of Health Services: Reflections on Practice." In *The Sage Handbook of Evaluation*, edited by I. F. Shaw, J. C. Green, and M. M. Mark, 461–485. London: Sage.

Lucas, R., E. F. Goldman, A. R. Scott, and V. Dandar. 2018. "Leadership Development Programs at Academic Health Centers: Results of a National Survey." *Academic Medicine* 93 (2): 229–236. doi:10.1097/ACM.0000000000001813.

Ludema, J. D., D. L. Cooperrider, and F. J. Barrett. 2006. "Appreciative Inquiry: The Power of the Unconditional Positive Question." In *Handbook of Action Research*, edited by P. Reason and H. Bradbury, 155–165. London: Sage.

Mabey, C. 2013. "Leadership Development in Organizations: Multiple Discourses and Diverse Practice." *International Journal of Management Reviews* 15 (4): 359–380. doi:10.1111/j.1468-2370.2012.00344.x.

MacDonald, B. 1993. "A Political Classification of Evaluation Studies in Education." In *Social Research: Philosophy, Politics and Practice*, edited by M. Hammersley, 105–109. London: SAGE. https://www.google.co.uk/books/edition/Social_Research/nMT7xt3CmgkC?hl=en&gbpv=1&printsec=frontcover

Maddox, J., and F. L. Replogle. 2019. "Strategic Change at a Hospital in Ecuador: Impact of Culture and Employee Engagement." *Organization Development Journal Fall* 37 (3): 21–30.

Mark, M. M., and R. L. Shortland. 1985. "Stakeholder-based Evaluation and Value Judgements." *Evaluation Review* 9 (5): 605–626. doi:10.1177/0193841X8500900504.

Martin, Graham P., David Kocman, Timothy Stephens, Carol J. Peden, and Rupert M. (2017). Pearse, "Pathways to professionalism? Quality improvement, care pathways, and the interplay of standardisation and clinical autonomy." *Sociology of Health & Illness* 39 (8): 1314–1329. https://onlinelibrary.wiley.com/doi/full/10.1111/1467-9566.12585

McCauley, C. D., and C. J. Palus. 2020. "Developing the Theory and Practice of Leadership Development: A Relational View." *The Leadership Quarterly* 32 (5): 101456. https://doi.org/10.1016/j.leaqua.2020.101456

McCray, J., Temple, P. F., & McGregor, S. (2021). Adult social care managers speak out: Exploring leadership development. *International Journal of Training and Development*, 25(2), 200–216.

McLean, S., and G. Moss. 2003. "They're Happy, but Did They Make a Difference? Applying Kirkpatrick's Framework to the Evaluation of a National Leadership Program." *Canadian Journal of Program Evaluation* 18 (1): 1–23.

Mertens, D. M. 2009. "*Transformative Research and Evaluation*". New York, NY: Guilford Press.

Mertens, D. M. 2015. "*Research and Evaluation in Education and Psychology: Integrating Diversity with Quantitative, Qualitative, and Mixed Methods.*" 4th ed. Thousand Oaks, CA: Sage publications.

Minshew, L. M., J. M. Zeeman, A. A. Olsen, A. A. Bush, J. H. Patterson, and J. E. McLaughlin. 2021. "Qualitative Evaluation of a Junior Faculty Team Mentoring Program." *American Journal of Pharmaceutical Education*. 85(4):288. Accessed 27 October 2021. Available at. doi:10.5688/ajpe8281.

Neubert, M. J., and L. D. Palmer. 2004. "Emergence of Women in Healthcare Leadership: Transforming the Impact of Gender Differences." *The Journal of Men's Health & Gender* 1 (4): 383–387. doi:10.1016/j.jmhg.2004.09.015.

Nuno-Solimis, R. 2017. "Revisiting Organizational Learning in Integrated Care." *International Journal of Integrated Care* 17 (4): 1–6.

O'Dwyer-Cunliffe, Finn., and Russell, Julian. (2020) "BAME representation and experience in the NHS." NHS Providers-Inclusive Leadership (blog). https://tinyurl.com/2muk726y

O'Sullivan, R. G. 2012. "Collaborative Evaluation within a Framework of Stakeholder-oriented Evaluation Approaches." *Evaluation and Program Planning* 35 (4): 518–522. doi:10.1016/j.evalprogplan.2011.12.005.

Okpala, P. 2020. "Increasing Access to Quality Healthcare through Collaborative Leadership." *International Journal of Healthcare Management* 13 (3): 229–230. doi:10.1080/20479700.2017.1401276.

Olano, B. L. 2015. "Preparing the New Nurse Leader for Health Care Delivery in South Africa in the Twenty-first Century." *African Journal for Physical, Health Education, Recreation and Dance* 1 (2): 485–497. *AJPHERD*) (Supplement

Pallesen, K.S., Rogers, L., Anjara, S., De Brún, A., and McAuliffe, E., 2020. A qualitative evaluation of participants' experiences of using co-design to develop a collective leadership educational intervention for health-care teams. Health expectations, 23(2): 358–367.

Patterson, T.E., Stawiski, S., Hannum, K.M. Chapmpion, and Down, H. (2017). Evaluating the impact of leadership development. 2nd Ed. 1–35. Greensboro, NC: Centre for Creative Leadership. https://eric.ed.gov/?id=ED167634 .

Patton, M. Q. 1994. "Developmental Evaluation." *Evaluation Practice* 15 (3): 311–319. doi:10.1177/109821409401500312.

Patton, M. Q. (1997). Utilization-focused evaluation: The new century text (3rd ed.). Sage Publications, Inc.

Patton, M. Q. 2008. *"Utilization-focused Evaluation: The New Century Text"*. 4th ed. Thousand Oaks: Sage Publications.

Patton, M. Q. 2015. *"Qualitative Research and Evaluation Methods: Integrating Theory and Practice"*. 4th ed. Thousand Oaks, CA: Sage Publications.

Pettigrew, A. M., R. W. Woodman, and K. S. Cameron. 2001. "Studying Organizational Change and Development: Challenges for Future Research." *Academy of Management Journal* 44 (4): 697–713.

Phillips, G., P. Lindeman, C. N. Adames, E. Bettin, C. Bayston, P. Stonehouse, D. Kern, A. K. Johnson, C. H. Brown, and G. J. Greene. 2019. "Empowerment Evaluation: A Case Study of Citywide Implementation within an HIV Prevention Context." *American Journal of Evaluation* 40 (3): 318–334. doi:10.1177/1098214018796991.

Phillips, J. J., and P. Phillips. 2007. "Measuring Return on Investment in Leadership Development." In *The Handbook of Leadership Development Evaluation*, edited by Hannum, K.M., Martineau, J.W. and Reinelt, C, 137–166. San Francisco, CA: John & Wiley Sons.

Picciotto, R. 2015. "Democratic Evaluation for the 21st Century." *Evaluation* 21 (2): 150–166. doi:10.1177/1356389015577511.

Plsek, P. E., and T. Wilson. 2001. "Complexity, Leadership, and Management in Healthcare Organisations." *British Medical Journal* 323 (7315): 746–749. doi:10.1136/bmj.323.7315.746.

Raju, P. G. 2021. "Healer Vs Leader-Determinants & Deterrents of Clinician Leadership in Indian Healthcare." *Indian Journal of Industrial Relations* 57 (2): 253–270.

Ramseur, P., M. A. Fuchs, P. Edwards, and J. Humphreys. 2018. "The Implementation of a Structured Nursing Leadership Development Program for Succession Planning in a Health System." *The Journal of Nursing Administration* 48 (1): 25–30. doi:10.1097/NNA.0000000000000566.

Rodriguez-Campos, L. 2012. "Stakeholder Involvement in Evaluation: Three Decades of the American Journal of Evaluation." *Journal of Multidisciplinary Evaluation* 8 (17): 57–79.

Russ-Eft, D., and H. Preskill. 2009. *Evaluation in Organizations: A Systematic Approach to Enhancing Learning, Performance, and Change*. 2nd ed. New York: Basic Books.

Schwartz, J., J. Bersin, and B. Pelster. 2014. *Human Capital Trends 2014 Survey: Top 10 Findings*. Deloitte Consulting LLP and Bersin by Deloitte. https://www2.deloitte.com/us/en/insights/focus/human-capital-trends/2014/human-capital-trends-2014-survey-top-10-findings.html

Scriven, M. 1991. *"Evaluation Thesaurus"*. 4th ed. Thousand Oaks, CA: Sage.

Scriven, M. 1999. "Pros and Cons about Goal-free Evaluation." *American Journal of Evaluation* 12 (1): 55–66.

Seigart, D. 2014. "Feminist Research Approaches to Studying School-based Health Care: A Three-country Comparison." In edited by Brisolara, S, Seigart, D. and SenGupta, S. *Feminist Evaluation and Research: Theory and Practice* 263–283. New York: The Guilford Press.

Sfantou, D. F., A. Laliotis, A. E. Patelarou, D. Sifaki-Pistolla, M. Matalliotakis, and E. Patelarou. 2017. "Importance of Leadership Style Towards Quality-of-care Measures in Healthcare Settings: A Systematic Review." *Healthcare*. 5 (4): 73. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/healthcare5040073

Sinclair, A. 2009. "Seducing Leadership: Stories of Leadership Development." *Gender, Work, and Organization* 16 (2): 266–284. doi:10.1111/j.1468-0432.2009.00441.x.

Smith, J., and J. Gosling. 2017. "Moving on Up: Women in Healthcare Leadership." *Practice Management* 27 (2): 24–27. doi:10.12968/prma.2017.27.2.24.

Solà, J. G., G. J. I Badia, P. D. M. Hito, M. A. C. Antonia Campo Osaba, and J. L. Del Val García. 2016. "Self-perception of Leadership Styles and Behaviour in Primary Health Care." *BMC Health Services Research* 16 (1): 572. doi:10.1186/s12913-016-1819-2.

Spencer, L., J. Ritchie, J. Lewis, and L. Dillon. 2003. *"Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence"*. London: CSRO.

Stacey, R. 2012. *Tools and Techniques of Leadership and Management: Meeting the Challenge of Complexity*. London: Routledge.

Stake, R. 1975. *"Program Evaluation, Particularly Responsive Evaluation"*. Kalamazoo, MI: Western Michigan University.

Stame, N. 2004. "Theory-based Evaluation and Varieties of Complexity." *Evaluation* 10 (1): 58–76. doi:10.1177/1356389004043135.

Strober, E. 2005. "Is Power-sharing Possible?" Using Empowerment Evaluation with Parents and Nurses in a Pediatric Hospital Transplantation Setting." *Human Organization* 64 (2): 201–210. doi:10.17730/humo.64.2.uh6u2exgheyxhxqj.

Stufflebeam, D. L. 1983. "The CIPP Model for Program Evaluation." In *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, edited by G. F. Madaus, M. Scriven, and D. L. Stufflebeam, 279–317. Boston: Kluwer: Nijhof Publishers.

Stufflebeam, D. L., and A. J. Shinkfield. 2007. *Evaluation Theory, Models, and Applications*. San Francisco, CA: Jossey-Bass.

Sucharew, H., and M. Macaluso. 2019. "Methods for Research Evidence Synthesis: The Scoping Review Approach." *Journal of Hospital Medicine* 14 (7): 416–418. doi:10.12788/jhm.3248.

Swanson, R.A., and Sleezer, C.M. (1987), "Training Effectiveness Evaluation", Journal of European Industrial Training 11 (4): 7–16. doi:10.1108/eb002227.

Tayabas, T. T., T. C. León, and J. M. Espino. 2014. "Qualitative Evaluation: A Critical and Interpretative Complementary Approach to Improve Health Programs and Services." *International Journal of Qualitative Studies on Health and Well-Being* 9 (1): 1–6.

Thiele, Graham, André Devaux, Claudio Velasco, and Douglas Horton. (2007). "Horizontal evaluation: fostering knowledge sharing and program improvement within a network." *American Journal of Evaluation* 28 (4): 493–508.

Tomlinson, M., D. O'Reilly, and M. Wallace. 2013. "Developing Leaders as Symbolic Violence: Reproducing Public Service Leadership through the (Misrecognized) Development of Leaders' Capitals." *Management Learning* 44 (1): 81–97. doi:10.1177/1350507612472151.

Turner, P. 2019. "Leadership Development Practices." In *Leadership in Healthcare*. Cham: Palgrave Macmillan. 295–324.

Wandersman, A. 2015. Getting to Outcomes: An Empowerment Evaluation Approach for Capacity Building and Accountability. *In* edited by Fetterman, D.M., Kaftrarian, S.J. and Wandersman, A. *Empowerment Evaluation: Knowledge and Tools for Self-assessment, Evaluation Capacity Building, and Accountability*. 2nd (ed.) 150–164. Sage Thousand Oaks, CA.

Wang, C., Y. L. Yuan, and M. L. Feng. 1996. "Photovoice as a Tool for Participatory Evaluation: The Community's View of Process and Impact." *Journal of Contemporary Health* 4 (3): 47–49.

Wäscher, S. S., S. Ritter, P. Vollman, J. Schildmann J, and J. Schildmann. 2017. "Methodological Reflections on the Contribution of Qualitative Research to the Evaluation of Clinical Ethics Support Services." *Bioethics* 31 (4): 237–245. doi:10.1111/bioe.12347.

Watkins, K. E., I. H. Lysø, and K. de Marrais. 2011. "Evaluating Executive Leadership Programs: A Theory of Change Approach." *Advances in Developing Human Resources* 13 (2): 208–239. doi:10.1177/1523422311415643.

Weiss, C. H. 1995. "Nothing as Practical as Good Theory: Exploring Theory-based Evaluation for Comprehensive Community Initiatives for Children and Families." In *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*, edited by P. Connell, A. C. Kubisch, L. B. Schorr, and C. H. Weiss. 1st ed., 65–92. Washington DC: Aspen Institute.

West, M. 2021. Compassionate Leadership: Sustaining wisdom, humanity and presence in health and social care. London: The Swirling Leaf Press.

West, M., K. Armit, L. Loewenthal, R. Eckert, T. West, and A. Lee. 2015. *Leadership and Leadership Development in Healthcare: The Evidence Base*. London: King's Fund.

Willmott, H. 1997. "Critical Management Learning." In *Management Learning: Integrating Perspectives in Theory and Practice*, edited by J. Burgoyne and M. Reynolds, 161–176. London: Sage.

World Health Organisation. 2020. "Global Strategy on Human Resources for Health." In *Workforce 2030: Health Workforce Department World Health Organization 20 Avenue Appia CH 1211*. Geneva, Switzerland. https://www.who.int/publications/i/item/9789241511131

Yang, T., M. G. Maa, Y. Lia, Y. Tiana, H. Liud, Y. Chene, S. Q. Zhangf, and J. Denga. 2020. "Do Job Stress, Health, and Presenteeism Differ between Chinese Healthcare Workers in Public and Private Hospitals: A Cross Sectional Study *Psychology*." *Health & Medicine* 25 (6): 653–666.