



Evaluation of Continuing Professional Development for Physicians – Time for Change: A Scoping Review

REVIEW

SHERA HOSSEINI

LOUISE ALLEN

FARAN KHALID

DONNY LI

ELIZABETH STELLRECHT

MICHELLE HOWARD

TERESA M. CHAN

*Author affiliations can be found in the back matter of this article

ubiquity press

ABSTRACT

Introduction: Evaluation of education interventions is essential for continuous improvement as it provides insights into how and why outcomes occur. Specifically, for physicians' continuing professional development (CPD) programs, which aim to upskill physicians in a range of practice-essential domains, evaluations are crucial to assure physicians' continuous development, enhanced patient care and safety. However, evaluations of health professions education (HPE) interventions tend to be outcomes focused, failing to capture how and why outcomes occur. This scoping review aimed to identify evaluation techniques used to evaluate CPD programs for physicians, and to determine how these techniques are being implemented as well as the their quality.

Methods: We searched PubMed, Embase, Web of Science, among others for English publications on evaluation of CPD programs for physicians, in the past decade. We used a data charting template to extract study details regarding the evaluation techniques and produced a checklist to assess the quality of the evaluations.

Results: 101 studies were included; of which 91 studies did not use an evaluation framework. Our findings revealed shortcomings in the evaluations of CPD programs including lack of attention to: intervention processes; unintended outcomes and contextual factors; use of theory; evaluation framework use; and rationale for chosen evaluation method.

Discussion: Our findings highlighted major gaps in the evaluation techniques employed in physicians' CPD. Attention needs to be paid to evaluating both program processes and outcomes to illuminate how and why impacts are or are not occurring.

CORRESPONDING AUTHOR:

Shera Hosseini, PhD

Postdoctoral Fellow, McMaster University, Faculty of Health Sciences, Department of Family Medicine, Canada

hosses@mcmaster.ca

TO CITE THIS ARTICLE:

Hosseini S, Allen L, Khalid F, Li D, Stellrecht E, Howard M, Chan TM. Evaluation of Continuing Professional Development for Physicians – Time for Change: A Scoping Review. *Perspectives on Medical Education*. 2023; 12(1): 198–207. DOI: <https://doi.org/10.5334/pme.838>

INTRODUCTION

Continuing professional development (CPD) refers to the activities that physicians undertake to maintain, update, and enhance their attitudes, knowledge, skills, performance, and relationships required to be effective clinicians, educators, researchers and leaders [1, 2]. The central goal of CPD is to enhance patient care and safety through the maintenance of evidence-based practice, the implementation of new clinical practices, and the “de-implementation” of non-evidence-based, outmoded practices [3]. Thus, evaluation of CPD is critical for determining the outcomes and effectiveness of various CPD activities as well understanding how and why certain outcomes occur.

In health professions education (HPE), educators mostly use outcome evaluation when evaluating interventions [4]. Outcome evaluation approaches look to answer the question ‘Did it work?’ focusing on the impacts of an intervention [4]. While it is important to understand the outcomes of an intervention, outcome evaluation approaches do not explain how and why outcomes are occurring, or what else is happening, as they fail to explore the processes that lead to outcomes, acknowledge the complexity of educational interventions, and consider unintended outcomes. Alternately, program evaluation refers to evaluation of design and implementation in addition to the outcomes and allows for understanding how and why programs work. There are a range of program evaluation models in HPE and healthcare broadly (e.g., realist evaluation; contribution analysis; context, input, process, product (CIPP) evaluation). Educators have called for a shift from an outcome evaluation approach to a program evaluation approach to shed light on the mechanistic processes of why and how programs work (or not) [5–7]. Additionally, it is unclear what program evaluation strategies are best applied to single CPD events or which might be better applied to an entire CPD providing unit (e.g., an Office of CPD).

Current reviews on CPD program effectiveness focus on reporting outcomes, with little consideration given to the implementation of evaluation models being used, and the quality of evaluations being conducted. For example, one review investigated the impacts of CPD with no attention to evaluation quality [8], while another focused on the use of outcome and program evaluation models and advocated for increased use of program evaluation models, as opposed to focusing on evaluation quality [4]. Thus, this study aims to conduct a scoping review to answer the following questions:

- What evaluation techniques are being used to evaluate CPD programs for physicians, and how are these techniques being implemented?
- What is the quality of evaluations of CPD programs for physicians?

METHODS

Our scoping review adhered to guidelines posed by Arksey and O’Malley [9].

SEARCH STRATEGY

A health sciences librarian (ES) conducted an electronic search with retrieval limited to English-language publications published from January 1, 2010 to March 24, 2021. We chose this time frame to include the most recent patterns of CPD program evaluation in an evolving field. ES searched PubMed, Embase via Elsevier, Web of Science Core Collection, Education Source via EBSCOhost, and APA PsycINFO via EBSCOhost using a combination of keywords and subject headings representing different medical specialties, continuing medical education, and program evaluation, ProQuest Dissertations and Theses Global was also searched. ES originally built the search in PubMed and translated it for the other databases (See Appendix 1 for the PubMed search). Citations were exported to Covidence for study management and screening [10].

STUDY SELECTION

Two reviewers (DL, FK) independently completed article screening against the study eligibility criteria at the title and abstract and performed full-text screening at the full-text level. Any disagreements were resolved through discussions. Studies that included an evaluation of structured CPD training or intervention (with various foci: e.g., teaching, mentorship, leadership, skill development) for physicians were included. Studies were excluded if they were: abstracts, commentaries, reviews, conference posters; included residents, undergraduate medical students, trainees, or any other non-MD participants in the CPD interventions; had no evaluation component; or had no CPD intervention.

DATA CHARTING

We utilized Microsoft Excel to develop a data charting template. The data charting template included study characteristics (e.g., country, setting, medical specialty, study design and structure, delivery mode, topic, use of theory in program development), and details of the evaluation (e.g., use of an evaluation model/framework, type of evaluation, and outcomes of the evaluation). DL and FK used this template to chart the data for all included studies. SH and LA then independently reviewed half of the charted necessary to assess author agreement and reach consensus.

COLLATING, SUMMARIZING, AND REPORTING THE DATA

Once data charting was complete, the study characteristics were quantified using quantitative content analysis. We

aimed to use a quality checklist to determine the quality of the included evaluations. However, upon exploration of the existing quality checklists, we discovered that while there are several with an educational focus, they focus on the overall study quality, as opposed to the quality of the evaluation conducted. We therefore developed a checklist to assess evaluation quality using a program evaluation lens. We used this lens to identify the gaps in current practice and highlight any particularly rigorous program evaluations. The checklist was developed by combining salient elements from the SQUIRE 2.0 Revised Standards for Quality Improvement Reporting Excellence [11], SQUIRE-EDU Standards for Quality Improvement Reporting Excellence in Education [12], and the TREND Statement checklist (for non-randomized evaluations of interventions) [13]. This was done as not all their components were relevant to evaluation quality. We chose these quality reporting tools as they are widely used. Two CPD experts reviewed the checklist and provided feedback, which was used to refine the checklist, but no items were added or removed.

The final checklist included 13 items, and for each item reviewers assigned it either a yes (met the criteria for the category), no (did not meet the criteria for the category), or partially (somewhat met the criteria for the item). These items were arranged across six categories of CPD intervention, methods (evaluation methodology, measures, analysis), results, and discussion.

For each item, along with its description, the checklist included the rationale for the item as well as the item source (e.g., adapted from an existing checklist, or author developed). For a detailed description of the checklist, refer to Appendix 2. SH and LA independently completed the

checklist for a quarter of the studies and then discussed decisions, including differences. Through this process additional notes were added to the checklist to help guide decision making with the aim of better aligning their completion of the checklist. The remaining studies were split evenly between SH and LA. Any uncertainties were resolved through discussion until consensus was reached. As this quality checklist aimed to explore what is being done well in program evaluation, and what needed improvement, rather than produce a score for each study, we narratively summarized the proportion of answers for each category.

EXPERT CONSULTATION

After initial analysis, expert reviewers reviewed the draft findings. The experts were key researchers in CPD and HPE. Seven experts reviewed an audio-recorded PowerPoint of this study’s rationale, methods, results, and discussion. The experts completed a 5-question Google form survey that inquired if the results aligned with what they would have expected, how to best present the findings, what to focus on in the discussion, and their thoughts on the most relevant elements of the findings to practice.

RESULTS

The searches returned 2152 articles; 101 studies were included (See Figure 1). More than half of the studies originated from North America, were in hospital, or community settings, and many had general practitioners as their primary participants (See Table 1). Studies were published across the entire study period with the most appearing in 2020 (n = 13).

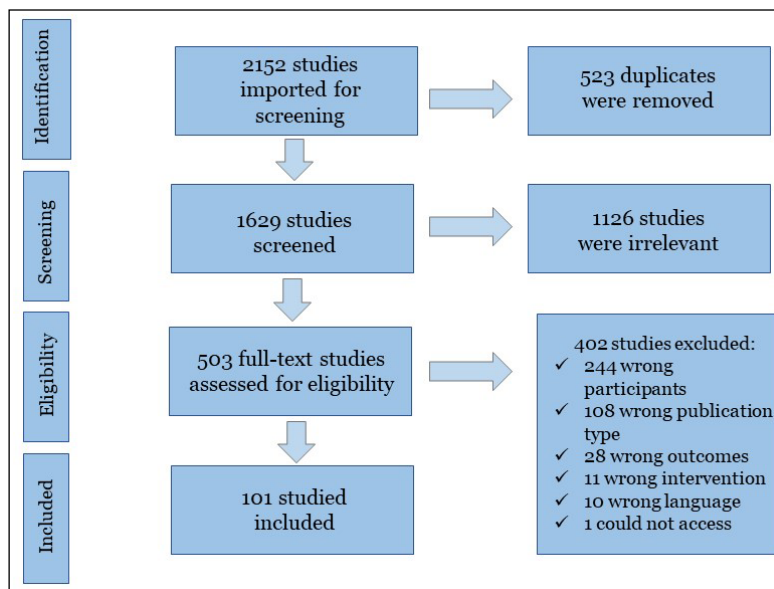


Figure 1 Flowchart of the literature search and study selection process.

STUDY CHARACTERISTIC	NUMBER OF STUDIES (%)
<i>Country of Study</i>	
USA	42 (41.6)
Canada	9 (8.9)
UK	8 (7.9)
Australia	5 (5.0)
Japan	4 (4.0)
China, Germany, Italy, Iran, Netherlands	3 (3.0)
Brazil, France, India, International	2 (2.0)
Ireland, Israel, Norway, Qatar, Saudi Arabia, South Africa, Spain, Sweden, Taiwan, Vietnam	1 (1.0)
<i>Setting</i>	
Hospital	47 (46.5)
Community	27 (26.7)
University	14 (13.9)
Other (nationwide, professional associations, private Organizations, Military, Not Clear)	11 (8.4)
<i>Specialty</i>	
GP/Primary Care Physician	38 (37.6)
Surgeon	16 (15.8)
Varied Specialties	14 (13.9)
Physician – not specified	7 (6.9)
Pediatrician	6 (5.9)
Internist	4 (4.0)
Hospitalist	3 (3.0)
Anesthesiologist, Oncologist, Psychiatrist, Radiologist	2 (2.0)
Cardiologist, Emergency Medicine, ICU, Pathologist, Rheumatologist	1 (1.0)

Table 1 Characteristics of included studies.

*Total percent greater than 100 as multiple options could exist per program.

The included studies included a variety of study designs with pre-post test designs (n = 68; 67.3%) most represented followed by post-test only (n = 25; 24.8%). See [Table 2](#) for study design characteristics.

The CPD programs included various foci (skill development, scholarship, teaching, leadership, mentoring, promoting well-being), types (interactive, didactic, mixed, other), modes of delivery (in class, online, hybrid), and number of session offerings (multiple, one-time) (See [Table 3](#)).

STUDY DESIGN CHARACTERISTIC	NUMBER OF STUDIES (%)
<i>Study Design</i>	
Pre-post test	68 (67.3)
Post-test only	25 (24.8)
Randomized Controlled Trial	4 (4.0)
Retrospective	2 (2.0)
Ethnographic, Prospective	1 (1.0)
<i>Method</i>	
Quantitative	58 (57.4)
Mixed methods	30 (29.7)
Qualitative	13 (12.9)
<i>Type of Measure*</i>	
Unvalidated survey	58 (57.4)
Interview	25 (24.8)
Record Review	19 (18.8)
Validated survey (reported survey validity data)	11 (10.9)
Performance Rating	11 (10.9)
Open-ended questions	10 (9.9)
Other	8 (7.9)
Focus Groups	7 (6.9)
Multiple Choice Questions- knowledge test questions	4 (4.0)
Patient Questionnaire	2 (2.0)
<i>Evaluation Framework</i>	
None specified	91 (90.1)
Kirkpatrick	7 (6.9)
Modified Kirkpatrick	2 (2.0)
Realist	1 (1.0)

Table 2 Design characteristics of included studies.

*Total percent greater than 100 as multiple options could exist per program.

EVALUATION QUALITY

We quantitatively summarized the frequencies of studies meeting each criterion across three categories (no, partial, yes). No study received a ‘yes’ for all the 14 criteria, with only one study [15] scoring ‘yes’ for 13 of the criteria. Most studies (n = 97, 96%) received a ‘yes’ for six or less of the criteria, with approximately half only receiving a ‘yes’ for one or less of the criteria (n = 45, 45%).

A few studies met only a limited number of the included criteria for evaluation techniques. For instance, most

CHARACTERISTIC	NUMBER OF STUDIES (%)
<i>Topic</i>	
Skill development	81 (80.2)
Scholarship	13 (12.9)
Teaching	6 (5.9)
Leadership	4 (4.0)
Mentoring	2 (2.0)
Well-being	1 (1.0)
<i>CPD Program Type</i>	
Mixed**	66 (65.3)
Interactive***	19 (18.8)
Didactic	11 (10.9)
Other	3 (3.0)
Not specified	2 (2.0)
<i>CPD Program Delivery*</i>	
In person	61 (60.4)
Hybrid	25 (24.8)
Online	16 (15.8)
Not reported	1 (1.0)
<i>CPD Program Number of Sessions</i>	
Multiple	60 (59.4)
One-off	29 (28.7)
Not specified	12 (11.9)

Table 3 Characteristics of Continuing Professional Development (CPD) programs.

*Total percent greater than 100 as multiple options could be selected.

**Mixed = combination of didactic and interactive learning.

***Interactive = learners actively involved in learning.

studies (81%) either did not use theories to elaborate on their CPD intervention design (n = 39) or had only explicated their program design in the context of current evidence (n = 44) without using a theory to explain why they believed their CPD intervention would work. Those that incorporated theory, referred to the following: the adult learning theory [16], theory of planned behavior [16], behavior change model [17], implementation change model [18], dual process theory [19], script theory [19], development process theoretical framework [20], social marketing theory [21], apprenticeship model [22], and the Comskil model [14].

The majority (88%) of the studies did not use an evaluation framework and among those that did, the Kirkpatrick framework was the dominant framework. Only one study used the realist framework to investigate the

intervention processes [15]. Further, 95% of evaluations did not provide a rationale for their evaluation method or justify why the selected evaluation approach would be appropriate based on their evaluation goals. The frequencies of studies meeting each criterion are illustrated in Table 4 (See Appendix 3 for Table 4).

EXPERT CONSULTATIONS

The consulted experts did not find the results unexpected, agreed with our presentation of results, and did not believe any key trends or literature were missing. Their input helped shape the focus of the discussion centering on how the methodological rigor of evaluations can be improved given the findings. Our experts emphasized the importance of highlighting the underutilization of theoretical underpinning and evaluation frameworks, insufficient elaboration on associations between intervention processes and outcomes, and absence of measures that would capture unintended outcomes.

DISCUSSION

This review sought to answer the following questions: a) *What evaluation techniques are being used to evaluate CPD programs for physicians, and how are these techniques being implemented?* and b) *What is the quality of evaluations of CPD programs for physicians?*

In terms of the evaluation techniques used, this review showed that few studies clearly articulated the frameworks that they utilized with only 10 studies clearly articulating the framework that they had used, and another 2 partially articulating this [23, 24]. This is problematic since research has specifically called for the use of evaluation frameworks in program evaluation efforts [5-7]. For evaluation implementation techniques and regardless of whether an evaluation framework was specified or not, outcome evaluation was far more common than more holistic program evaluation. This aligns with recent research that shows evaluation in HPE is largely outcome evaluation [4], however, outcome-oriented approaches can potentially fail to provide a holistic presentation of the mechanistic processes that may have contributed to the occurrence of the outcomes or lack thereof. Outcome evaluations tend to focus on evaluation approaches with linear causal relationship assumptions about the program elements and outcomes and overlook program's theory and the complexity of the educational interventions. In turn, such evaluations may only provide information regarding the educational interventions' outcomes but lack the ability to inform the processes through which the desired or unintended outcomes occur and why these outcomes are or are not observed [25, 26].

Research has both demonstrated and underlined the use of theory (e.g., social theory of learning) in elucidating the potential impacts associated with CPD programs [27]. In particular, the use of complexity and system theories has been encouraged due to the complex interplay of multiple interacting components as well as the non-linear relationships between the intervention elements and outcomes in HPE [25]. Educators and evaluators need to pay close attention to the theories underpinning evaluation frameworks and whether they align with their specific evaluation objectives when determining which evaluation framework to employ. The use of theory-driven evaluations, such as the realist framework, is encouraged in conjunction with more outcome-focused evaluation models since theory-based evaluations allow for elucidating the mechanisms that connect program processes to its outcomes and why certain outcomes may or may not be occurring.

Implementation science researchers have recognized the need for determining interventions' theoretical bases that would assist with understanding the mechanisms of change at play [28–30]. This work can help inform evaluation in HPE. For example, Nilsen (2015) introduced a framework that entailed three main aims for using theoretical approaches that are used in implementing interventions: *a) describing / guiding the process of translating research into practice; b) understanding / explaining what influences implementation outcomes; and c) evaluating implementation*. These three aims were further broken down to five categories of theories, models, and frameworks: *process models, determinant frameworks, classic theories, implementation theories, and evaluation frameworks*. The authors asserted several advantages associated with the use of formal theories in developing interventions including the ability to openly question, challenge, refine, adapt, or refute theories; a greater degree of consistency between theories and the existing facts; and the ability of theories to offer meaningful contexts to facts, promoting the construction of a unified body of knowledge [31].

In terms of the quality of CPD evaluations, we developed and utilized a quality checklist to determine the quality of the evaluations conducted. We found that when using a program evaluation lens to assess the quality of evaluations, the evaluation quality for most studies was low. This was true regardless of study type, while the mean quality checklist score was higher for qualitative studies (mean = 14), than quantitative studies (mean = 7) and mixed method studies (mean = 10), the mean scores were at most just over half of the maximum score. This was unsurprising, as quantitative studies often focus on outcomes only, while half of the mixed methods studies used less in-depth methods of qualitative data collection such as open-ended questions and record review.

Assessing the studies against the checklist allowed us to depict a better presentation of the current state of the literature, highlight gaps, underline exemplar CPD evaluation approaches, and offer future recommendations. Prominent gaps pertaining to the conduct and reporting of CPD evaluations included: inadequate attention paid to theory in intervention design; unclear descriptions of evaluation models; emphasis on the categorization of program outcome evaluation and minimal utilization of evaluation models that include more than simple classification of outcomes; infrequent reporting of rationale for choosing specific methods for analysis; limited reporting of unintended consequences and links between the interventions and contextual factors; as well as insufficient attention paid to the interventions' causal pathways and interpretation of the findings in the context of current evidence and theory. Table 5 offers some considerations on how to address these gaps with related rationale. Table 5 also illustrates an example using an exemplar study that received the highest score from our checklist, to illustrate how these considerations can be done in practice.

This review revealed considerable opportunities for improving the quality of CPD evaluations. One reason for less attention paid to these elements when conducting CPD evaluations might be that it is only recently HPE journals are recognizing and requiring inclusion of theory in the submissions they publish. Furthermore, recognition of contextual elements in evaluations is another recent phenomenon particularly in CPD where few studies have used approaches that help understand how and why impacts may be occurring rather than merely 'did impacts occur?' Moreover, compared to Undergraduate Medical Education (UME), a great number of CPD offices tend to have less access to doctoral-level trained educators and many working in this space may not have the required training to undertake more rigorous evaluations. We also imagine that there may be a lag in the uptake of key reporting guidelines like the SQUIRE-EDU extension that could help improve outcomes reporting when evaluating CPD.

Although access to limited resources and evaluation expertise may contribute to the lack of evaluation framework use, and the use of simple taxonomies of program outcomes when frameworks are used, it is important to note that the short-term effects such as increased knowledge do not necessarily result in more long-term outcomes such as practice changes, enhanced care quality or patient-related outcomes [32, 33]. More long-term program evaluations that pay simultaneous attention to both processes (including contextual/external elements) and impacts (including unintended outcomes) of the CPD interventions and incorporate the use of theory using program evaluation models such as the logic model,

CONSIDERATION 1 – CONSIDER WHY YOU HAVE CHOSEN A PARTICULAR EVALUATION MODEL AND BE EXPLICIT WITH YOUR RATIONALE.

Importance: Justifying the use of a particular evaluation model allows evaluators to examine if the model is best placed to help understand how and why outcomes are occurring. Communicating this in published evaluations gives transparency to the evaluation methods and allows readers to gain insight into why a model was chosen, helping them to make a judgment on the appropriateness of the evaluation.

Example: McDaniel and colleagues utilize realist evaluation to evaluate a clinical faculty mentorship program. Realist evaluation is one of many types of program evaluation that allows for exploration of outcomes and helps to explain how and why they occur. They explain what realist evaluation is, why they took a qualitative approach as a result of using realist evaluation and the three-phase realist evaluation process they employed. They also explain the underlying concept of realist evaluation which focuses on context, mechanisms, and outcomes. Their full explanation can be found on page 105 under the conceptual framework section of the methods [15].

CONSIDERATION 2 – BE CLEAR ON WHY THE CPD PROGRAM IS EXPECTED TO WORK (E.G., PROGRAM THEORY).

Importance: Program theory is central to evaluation. Developing program theory should be part of the development of educational intervention. It is often developed and refined in an iterative process. Developing and communicating program theory allows evaluators (and readers of published evaluations) to understand how and why the education intervention is expected to work. Not only does it provide a foundation for conducting the evaluation, but also helps those wishing to implement a similar intervention understand how the program is leading to the outcomes.

Example – McDaniel and colleagues present a program theory that draws on contexts, mechanisms, and outcomes. This is in line with their use of a realist evaluation model, as mentioned in consideration 1. While the program theory is iterative, articulating the initial program theory guides both the development of the educational intervention and the evaluation. It also provides the readers insight into why and how the authors thought the educational intervention would work. The program theory is on page 105–106 [15].

CONSIDERATION 3 – MEASURE PROCESSES AS WELL AS OUTCOMES, TAKING A BROAD VIEW OF OUTCOMES CONSIDERING UNINTENDED OUTCOMES AND CONTEXTUAL AND EXTERNAL FACTORS TO HELP UNDERSTAND HOW AND WHY OUTCOMES ARE OCCURRING.

Importance: Capturing information about the development and implementation of the intervention (processes) is crucial to understanding how and why the program outcomes did or did not occur, as well as why any unintended outcomes might have resulted. This information allows us to understand if the intervention was implemented as intended and what else was happening that may have contributed to the outcomes, or that might be a reason for why outcomes were not observed. Additionally, educational interventions do not occur in a bubble, therefore contextual factors such as resources, culture, and level of support (just to name a few) can affect whether outcomes do or do not occur, and as such it is important to take these into consideration. The collection of this information can be informed by program theory and program evaluation models.

Example: The McDaniel and colleagues’ paper would have benefited from exploration of the processes of the intervention – that it was the mentorship program implemented as intended – however they do describe multiple barriers to the program which speaks to implementation, and why some may not have experienced particular outcomes (page 110). They do a good job of capturing both unintended outcomes, as well as contextual and external factors that help understand how the outcomes occurred. This is because realist evaluation explicitly calls out the importance of context as a key element of the program theory. They identified four context domains: (1) past personal experience, (2) current competing priorities, (3) institutional culture, and (4) gaps in support and resources that influenced the outcomes. They also identified three outcomes that were not the same as those articulated in the initial program theory. For a full description of the contexts and outcomes see the results section of the paper, pages 106–110 [15].

CONSIDERATION 4 – LINK RESULTS BACK TO PROGRAM THEORY AND DESCRIBE MECHANISMS FOR HOW AND WHY INTERVENTIONS DID OR DID NOT WORK.

Importance: The results should be linked back to the program theory, and consideration should be given to whether the initial program theory holds, or whether revision of the theory is required. Results should be discussed in the context of the program theory, and they should elaborate on potential reasons as to why certain outcomes did or did not occur. This allows consideration of if it was the intervention leading to the outcomes, or if there were other factors at play.

Example: McDaniel and colleagues address the mechanisms for how and why the program worked, or didn’t work in the mechanisms, barriers, and revised program theory sections of the results. The mechanisms include connecting with faculty, sharing ideas and strategies and self-reflecting. And the barriers include time and location limitations, perceived lack of fit, and individual priorities. While the revised program theory focuses on positive outcomes, the presentation of the barriers helps to understand why some participants of the program may not have experienced these outcomes. For full details see pages 108–110 [15].

Table 5 Considerations on how to improve evaluation in continuing professional development (CPD).

realist evaluation, CIPP, contribution analysis, layered analysis, among several others, are needed.

Moreover, in line with enhancing the quality of CPD evaluations, evaluation utility is an important point worth emphasizing. Keeping evaluation use at the forefront of the intervention design, development, and evaluation from program conception and design to evaluation will contribute to the evaluation quality and relevance for use by stakeholders. If users are to consider the evaluation

findings, evaluation utility would need to accompany every step of the evaluation process [34, 35]. Lastly, articulation of causal mechanisms and processes elucidating how and why programs work or not is critical and needs to be undertaken using theory-driven evaluations [36–38]. CPD programs are a significant component of the complex HPE pedagogical interventions that if evaluated properly, would allow for further reform and positive change. In doing so, undertaking rigorous evaluations with strong theoretical

underpinning and use of evaluation frameworks that pay close attention to these contextual and mechanistic complexities is crucial.

LIMITATIONS

The present review included focusing on articles published after 2010. Nevertheless, this cut-off was selected to be current with the topical trends of CPD program evaluation. Further, this review may have overlooked those studies that did not specify their interventions as CPD, or those not published in English. Finally, although it would be interesting to see a comparison between MDs and non-MDs, this went beyond the scope of our review, which was originally conceived to include MDs only.

CONCLUSION

This scoping review examined the span and quality of CPD evaluation techniques, and their methods of implementation in CPD programs for physicians. We developed a qualitative checklist using existing prominent checklists, that could further enhance the quality of future CPD program evaluations, which may guide studies that examine CPD program development and evaluation and highlight existing gaps associated with the evaluation of CPD programs. Salient areas in need of deliberation highlight considerations of broader and more long-term impacts of the interventions and their processes, use of theory and evidence when designing and evaluating the CPD programs, adherence to evaluation frameworks, augmenting quantitative approaches with qualitative measures to detect unintended outcomes, and lastly reflection on the potential causal pathways through which the programs may exert their impacts.

ADDITIONAL FILES

The additional files for this article can be found as follows:

- **Supplemental File 1.** List of included studies. DOI: <https://doi.org/10.5334/pme.838.s1>
- **Appendix 1.** PubMed Search. DOI: <https://doi.org/10.5334/pme.838.s2>
- **Appendix 2.** Qualitative Checklist. DOI: <https://doi.org/10.5334/pme.838.s3>
- **Appendix 3.** Table 4. Checklist for quality appraisal of Continuing Professional Development (CPD) program evaluations in individual studies. DOI: <https://doi.org/10.5334/pme.838.s4>

ACKNOWLEDGEMENTS

We would like to thank our expert panel for offering invaluable insights regarding how to communicate our review findings more effectively.

COMPETING INTERESTS

The authors have no competing interests to declare.


FUNDING INFORMATION

This study was supported by a Postdoctoral Fellowship from MIRA and MERIT at McMaster University.

AUTHOR AFFILIATIONS

Shera Hosseini, PhD  orcid.org/0000-0002-7144-8510

Postdoctoral Fellow, McMaster University, Faculty of Health Sciences, Department of Family Medicine, Canada

Louise Allen, PhD, GradCertEdDes, BNutrDiet (Hons)  orcid.org/0000-0003-1087-2235

Lecturer, Monash Center for Professional Development and Monash Online Education, Australia

Faran Khalid, BSc  orcid.org/0000-0001-5410-1729

Medical Student, Michael G. DeGroote School of Medicine, Hamilton, Ontario, Canada

Donny Li  orcid.org/0000-0003-2746-6794


Undergraduate Student, Faculty of Health Sciences, McMaster University, Hamilton, Canada; Research Student, Division of Supportive and Palliative Care, William Osler Health System, Brampton, Canada; Research Student, Department of Research, Humber River Hospital, Toronto, Canada

Elizabeth Stellrecht, MLS  orcid.org/0000-0002-9524-6171

Associate Librarian, Head of Health Sciences Library Services, University Libraries, University at Buffalo, Buffalo, NY, US

Michelle Howard, PhD  orcid.org/0000-0001-8127-5492

Associate Professor, Department of Family Medicine, McMaster University, 100 Main S. W., Hamilton ON, L8P 1H6, Canada

Teresa M. Chan, MD, FRCPC, MHPE[CT1]  orcid.org/0000-0001-6104-462X

Associate Professor, Division of Emergency Medicine, Department of Medicine, McMaster University, Hamilton, Canada; Associate Dean, Continuing Professional Development, Faculty of Health Sciences, McMaster University, Canada; Clinician Scientist, McMaster Education Research, Innovation, and Theory (MERIT) program, Faculty of Health Sciences, McMaster University Hamilton, Canada

REFERENCES

- Filipe HP, Silva ED, Stulting AA, Golnik KC.** Continuing professional development: Best Practices. *Middle East Afr J Ophthalmol.* 2014; 21(2): 134. DOI: <https://doi.org/10.4103/0974-9233.129760>
- Sargeant J, Bruce D, Campbell CM.** Practicing physicians' needs for assessment and feedback as part of professional development. *J Contin Educ Health Prof.* 2013; 33(S1). DOI: <https://doi.org/10.1002/chp.21202>
- Prasad V, Vandross A, Toomey C, Cheung M, Rho J, Quinn S,** et al. A decade of reversal: an analysis of 146 contradicted medical practices. *Mayo Clin Proc.* 2013; 88(8): 790–8. DOI: <https://doi.org/10.1016/j.mayocp.2013.05.012>
- Allen LM, Hay M, Palermo C.** Evaluation in health professions education—Is measuring outcomes enough? *Med Educ.* 2022; 56(1): 127–36. DOI: <https://doi.org/10.1111/medu.14654>
- Yardley S, Dornan T.** Kirkpatrick's levels and education 'evidence'. *Med Educ.* 2012; 46(1): 97–106. DOI: <https://doi.org/10.1111/j.1365-2923.2011.04076.x>
- Parker K.** A better hammer in a better toolbox: Considerations for the future of programme evaluation. *Med Educ.* 2013; 47(5): 440–2. DOI: <https://doi.org/10.1111/medu.12185>
- Haji F, Morin MP, Parker K.** Rethinking programme evaluation in health professions education: beyond 'did it work?'. *Med Educ.* 2013; 47(4): 342–51. DOI: <https://doi.org/10.1111/medu.12091>
- Allen LM, Palermo C, Armstrong E, Hay M.** Categorizing the broad impacts of continuing professional development: a scoping review. *Med Educ.* 2019; 53: 1087–99. DOI: <https://doi.org/10.1111/medu.13922>
- Arksey H, O'Malley L.** Scoping studies: towards a methodological framework. *Int J Soc Res.* 2005; 8(1): 19–32. DOI: <https://doi.org/10.1080/1364557032000119616>
- Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. www.covidence.org.
- Goodman D, Ogrinc G, Davies L, Baker GR, Barnsteiner J, Foster TC,** et al. Explanation and elaboration of the SQUIRE (Standards for Quality Improvement Reporting Excellence) Guidelines, V.2.0: examples of SQUIRE elements in the healthcare improvement literature. *BMJ Qual Saf.* 2016; 1–24. DOI: <https://doi.org/10.1136/bmjqs-2015-004480>
- Ogrinc G, Armstrong GE, Dolansky MA, Singh MK, Davies L.** SQUIRE-EDU (Standards for Quality Improvement Reporting Excellence in Education): Publication Guidelines for Educational Improvement. *Acad Med.* 2019; 24(10): 1461–70. DOI: <https://doi.org/10.1097/ACM.0000000000002750>
- Des Jarlais DC, Lyles C, Crepaz N, TREND Group.** Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *Am J Public Health.* 2004; 94: 361–6. DOI: <https://doi.org/10.2105/AJPH.94.3.361>
- Bylund CL, Alyafei K, Afana A, Al-Romaihi S, Yassin M, Elnashar M,** et al. Satisfaction with a 2-day communication skills course culturally tailored for medical specialists in Qatar. *J Family Community Med.* 2017; 24(2): 122–7. DOI: <https://doi.org/10.4103/2230-8229.205118>
- McDaniel CE, Rooholamini SN, Desai AD, Reddy S, Marshall SG.** A Qualitative Evaluation of a Clinical Faculty Mentorship Program Using a Realist Evaluation Approach. *Acad Pediatr.* 2020; 20(1): 104–12. DOI: <https://doi.org/10.1016/j.acap.2019.08.008>
- Adams RE, Laraque D, Chemtob CM, Jensen PS, Boscarino JA.** Does a one-day educational training session influence primary care pediatricians' mental health practice procedures in response to a community disaster? Results from the reaching children initiative (RCI). *Int J Emerg Ment Health.* 2013; 15(1): 3–14.
- Pimenta HB, Caldeira AP, Mamede S.** Effects of 2 educational interventions on the management of hypertensive patients in primary health care. *J Contin Educ Health Prof.* 2014; 34(4): 243–51. DOI: <https://doi.org/10.1002/chp.21252>
- Porcheret M, Main C, Croft P, Dziedzic K.** Enhancing delivery of osteoarthritis care in the general practice consultation: evaluation of a behaviour change intervention. *BMC Fam Pract.* 2018; 19(1): 26. DOI: <https://doi.org/10.1186/s12875-018-0715-8>
- Schaye V, Janjigian M, Hauck K, Shapiro N, Becker D, Lusk P,** et al. A workshop to train medicine faculty to teach clinical reasoning. *Diagnosis (Berl).* 2019; 6(2): 109–13. DOI: <https://doi.org/10.1515/dx-2018-0059>
- Shah P, Cross V, Sii F.** Sailing a safe ship: improving patient safety by enhancing the leadership skills of new consultant specialist surgeons. *J Contin Educ Health Prof.* 2013; 33(3): 190–200. DOI: <https://doi.org/10.1002/chp.21184>
- Vasudev K, Lamoure J, Beyaert M, Dua V, Dixon D, Eadie J,** et al. Academic detailing among psychiatrists – feasibility and acceptability. *Int J Health Care Qual Assur.* 2017; 30(1): 79–88. DOI: <https://doi.org/10.1108/IJHCQA-04-2016-0047>
- Zhang CY, Fang F, Peng MQ, Zhao Y, Liu RX, Jia CB.** Qualitative evaluation of the general practitioner chronic non-communicable diseases training programme. *BMC Med Educ.* 2020; 20(1): 7. DOI: <https://doi.org/10.1186/s12909-020-02226-x>
- Balmer DF, Richards BF.** Faculty Development as Transformation: Lessons Learned From a Process-Oriented Program. *Teach Learn Med.* 2012; 24(3): 242–7. DOI: <https://doi.org/10.1080/10401334.2012.692275>
- Pluye P, Grad R, Granikov V, Theriault G, Fremont P, Burnand B,** et al. Feasibility of a knowledge translation CME program: Courriels Cochrane. *J Contin Educ Health Prof.* 2012; 32(2): 134–41. DOI: <https://doi.org/10.1002/chp.21136>

25. **Frye AW, Hemmer PA.** Program evaluation models and related theories: AMEE Guide No. 67. *Med Teach*. 2012; 34(5): e288–e99. DOI: <https://doi.org/10.3109/0142159X.2012.668637>
26. **Parker K, Burrows G, Nash H, Rosenblum ND.** Going beyond Kirkpatrick in evaluating a clinician scientist program: it's not if it works but how it works". *Acad Med*. 2011; 86(11): 1389–96. DOI: <https://doi.org/10.1097/ACM.0b013e31823053f3>
27. **Allen LM, Hay M, Armstrong E, Palermo C.** Applying a social theory of learning to explain the possible impacts of continuing professional development (CPD) programs. *Med Teach*. 2020; 10(42): 1140–47. DOI: <https://doi.org/10.1080/0142159X.2020.1795097>
28. **Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A.** Making psychological theory useful for implementing evidence based practice: a consensus approach. *Qual Saf Health Care*. 2005; 14: 26–33. DOI: <https://doi.org/10.1136/qshc.2004.011155>
29. **Sales A, Smith J, Curran G, Kochevar L.** Models, strategies, and tools: theory in implementing evidence-based findings into health care practice. *J Gen Intern Med*. 2006; 21: S43–9. DOI: <https://doi.org/10.1111/j.1525-1497.2006.00362.x>
30. **Kitson AL, Harvey G, McCormack B.** Enabling the implementation of evidence-based practice: a conceptual framework. *Qual Health Care*. 1998; 7: 149–58. DOI: <https://doi.org/10.1136/qshc.7.3.149>
31. **Nilsen P.** Making sense of implementation theories, models and frameworks. *Implement Sci*. 2015; 10(53). DOI: <https://doi.org/10.1186/s13012-015-0242-0>
32. **Davis D, Thomson O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A.** Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? *JAMA*. 1999; 282: 867–74. DOI: <https://doi.org/10.1001/jama.282.9.867>
33. **Casper ES.** Using implementation intentions to teach practitioners: changing practice behaviors via continuing education. *Psychiatr Serv*. 2008; 59(7): 747–52. DOI: <https://doi.org/10.1176/ps.2008.59.7.747>
34. **Shadish WR, Cook TD, Leviton L.** *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, CA: Sage; 1991.
35. **Patton MQ.** *Utilization-Focused Evaluation* (4th ed). Thousand Oaks, CA: Sage; 2008.
36. **Chen H.** *Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective* (2nd ed.). Thousand Oaks, CA: Sage; 2015. DOI: <https://doi.org/10.4135/9781071909850>
37. **Weiss CH.** *Evaluation: Methods for studying programs and policies* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall; 1998.
38. **Weiss CH.** Which links in which theories shall we evaluate? In A Petrosino, Rogers PJ, Huebner TA, Hacsí TA (eds.), *Program Theory in Evaluation: Challenges and Opportunities*. *New Dir Eval*. 2000; 87: 35–45. DOI: <https://doi.org/10.1002/ev.1180>

TO CITE THIS ARTICLE:

Hosseini S, Allen L, Khalid F, Li D, Stellrecht E, Howard M, Chan TM. Evaluation of Continuing Professional Development for Physicians – Time for Change: A Scoping Review. *Perspectives on Medical Education*. 2023; 12(1): 198–207. DOI: <https://doi.org/10.5334/pme.838>

Submitted: 08 December 2022 **Accepted:** 12 May 2023 **Published:** 02 June 2023

COPYRIGHT:

© 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Perspectives on Medical Education is a peer-reviewed open access journal published by Ubiquity Press.