

RESEARCH

Open Access



A qualitative study of evaluators of Canadian postgraduate medical education programs: challenges and recommendations

Jenna D. Milosek^{1*} and Katherine A. Moreau^{1,2}

Abstract

Background In Canada, postgraduate medical education has transitioned to Competence by Design, a form of competency-based medical education. Program evaluation can help us learn more about the efficiency and effectiveness of Competence by Design and the transition to it. While most Canadian postgraduate medical education programs conduct program evaluations to support improvement and accreditation, the quality of these evaluation activities is questionable. Program evaluation within these contexts is challenging due to limited time, a lack of knowledge of program evaluation, and limited resources. The purpose of this study is to investigate how evaluators support postgraduate medical education programs and how they can help build the capacity of these programs to conduct evaluations and utilize the findings.

Methods Using a qualitative study design, we employed purposeful sampling to interview evaluators with knowledge and experience in evaluating postgraduate medical education programs. We conducted interviews with 12 evaluators. We thematically coded the data using a deductive and inductive process.

Results Evaluators can support postgraduate medical education programs as temporary and external consultants or as internal evaluation experts employed by the universities where the programs reside. Evaluators can help programs build their capacities to do program evaluation and use the findings by suggesting evaluation approaches and methods, leveraging existing data, interpreting the findings, encouraging programs to share conclusions, co-developing a manageable follow-up plan, and establishing quality assurance that supports accreditation.

Conclusions This study presents empirical research on program evaluation in medical education, as well as information for evaluators embarking on evaluations of medical education programs. While this study confirmed that some programs leverage qualified evaluators, the limited internal evaluation units within Canadian medical education universities are concerning. If the potential of evaluation is to be realized and sustainable in medical education, allocating adequate resources is essential. Selecting the appropriate evaluation type, questions, and approaches can support program directors and evaluators in continuing to transform medical education.

Keywords Program evaluation to support accreditation, Postgraduate medical education

*Correspondence:

Jenna D. Milosek
jmilosek@uottawa.ca

¹Faculty of Education, University of Ottawa, 145 Jean-Jacques-Lussier Private, Ottawa, ON K1N 9A8, Canada

²Centre for Research On Educational and Community Services, University of Ottawa, 136 Jean-Jacques-Lussier Private #5002, Ottawa, ON K1N 9A8, Canada



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Between 2017 and 2025, postgraduate medical education (PGME) programs in Canada transitioned to Competence by Design (CBD), the national model for competency-based medical education [1, 2]. CBD combines a time- and outcomes-based approach to the design, implementation, assessment, and evaluation of specialty residency training programs, using the CanMEDS sequenced competencies framework [3]. CBD is a transformational, multi-year initiative in which residents progress through the training stages using experience-based learning and validated competency assessments [3].

Given the scale of these changes and the resources involved, program evaluation is essential to ensure accountability and guide improvements in these programs [4–6]. Instituting CBD across Canadian PGME programs is relevant to evaluation because evaluation results can demonstrate that institutions are pursuing highly desired learning programs that train higher-quality residents for practice. Evaluations that incorporate the perspectives of faculty, residents, and patients can foster learning and responsiveness to implementation challenges [5, 7–9]. Ongoing evaluation of CBD is crucial. It identifies areas for refinement, measures program outcomes, and contributes to the overall effectiveness of PGME programs [5, 10–12]. Furthermore, evaluating CBD-based programs differs from evaluating other training programs, as the organization of relevant competencies creates opportunities to evaluate PGME programs that may not exist elsewhere. Additionally, PGME programs are complex learning contexts and, despite their similarities, what works in one specialty program may not be applicable in another [13]. In theory, CBD initiatives should be evaluated by examining program fidelity before assessing outcomes, to ensure that the program's theory has been appropriately implemented and will eventually lead to the anticipated results [14].

Although most PGME programs conduct evaluations for quality improvement and accreditation, concerns remain about the quality and sustainability of these efforts. Many programs face significant barriers, including limited time, a lack of expertise or personnel, and insufficient funding for evaluation activities [13, 15, 16]. This raises a critical question: Should program directors, often lacking formal training in evaluation, be expected to lead these efforts, or would dedicated evaluators be better positioned to support this work? [13, 17–19].

Evaluators are vital contributors to the evaluation processes. They help empower stakeholders, generate credible evidence to inform policy and practice, and ensure that improvements are intentional [20–23]. Using established standards focused on utility, feasibility, propriety, and accuracy, evaluators bring methodological rigour and ethical grounding to the processes [24–26]. In

addition to conducting evaluations, evaluators can play a critical role in evaluation capacity building (ECB), which strengthens individuals and organizations to embed evaluation into routine practices. Through ECB, evaluators can help PGME programs and their stakeholders (e.g., residents, staff physicians, administrative staff, patients) develop the skills and structures necessary to plan and conduct rigorous evaluations and use the results [27–30]. This capacity can be built through formal training (e.g., workshops, collaboration with experienced evaluators) or informal learning (e.g., participation in evaluation activities), enhancing stakeholders' confidence and competence [31].

Moreover, evaluators can foster the use of evaluation by promoting evaluation literacy, supporting informed decision-making, and demonstrating the learning benefits of applying evaluation findings [32, 33]. While evaluation findings often drive program improvement, resistance can occur due to concerns about scrutiny, resource limitations, or organizational inertia, particularly when findings highlight problems [31]. To mitigate such risks and ensure evaluation results are trusted and actionable, evaluators must apply methodologically sound approaches and engage stakeholders throughout the processes [34–36].

To better understand and improve the role of program evaluation in PGME programs, this study examined how evaluators support PGME programs and how they can help build the capacity of these programs to conduct evaluations and utilize the findings. Specifically, the following questions guided this study:

1. How are evaluators supporting PGME programs with program evaluation?
2. How can evaluators help PGME program stakeholders build their capacities *to do* program evaluation?
3. How can evaluators help PGME program stakeholders develop their capacities *to use* evaluation findings?

Methods

Paradigmatic orientation

We applied pragmatism as our paradigmatic orientation to this study. We emphasized the research problem and employed all available approaches to better understand it [37]. We are concerned with applications of what works and solutions to problems [38].

Sample

We employed purposeful sampling to recruit evaluators with relevant expertise in evaluating PGME programs in Canada [39]. We used two recruitment strategies. First, we emailed 33 evaluators affiliated with the

American Evaluation Association (AEA)'s *Health Evaluation* or *Health Professions Education, Evaluation, and Research* Topical Interest Groups (TIGs). Second, we emailed 19 evaluators associated with Canadian English-language PGME programs and not affiliated with the TIGs mentioned above. Additionally, follow-up emails were sent to evaluators who did not respond to the first email.

Interview guide development

To inform the development of our interview guide, we drew on literature about evaluators [18, 19, 40, 41] and evaluation capacity building [27–31, 36, 42, 43]. The interview guide used is included in the Supplementary Information Section. After piloting the interview guide and incorporating feedback, the final guide consisted of 17 questions focusing on three core areas: (a) how evaluators currently support PGME programs, (b) how evaluators can help programs build capacity to conduct evaluations, and (c) how evaluators can help programs build capacity to use the evaluation findings [44].

Data collection

We emailed the 52 eligible evaluators a letter of information and a consent form and asked them to reply if they were interested in participating in the study. Before each one-hour Zoom interview, participants provided written informed consent. Interviews were recorded and transcribed verbatim. Each participant received a \$10 Amazon gift card.

Data analysis

Using a cyclical process of data condensation, data display, and drawing/verifying conclusions, we analyzed the data by hand [45]. Following each interview, we listened to the recording collectively and created written memos to summarize the content, identify themes, and outline probes for subsequent interviews. During data condensation, we independently reviewed the memos, read the verbatim transcripts, and inductively coded the data. We then met to compare, refine, and establish a shared coding scheme that incorporated both inductive insights and, where relevant, deductively derived codes that informed the evaluation literature and our research questions. In the data display phase, we organized codes into tables and diagrams to visualize relationships across themes. Throughout, we maintained detailed memos to document our coding decisions and sought feedback from two other program evaluation researchers. Once we reached consensus, we collaboratively synthesized the coded data, selected exemplar quotations, and finalized our interpretations. Our use of an audit trail, multiple coders, and feedback mechanisms enhanced the trustworthiness of the findings.

Table 1 The barriers and facilitators to evaluation and evaluators' solutions/strategies

		Solutions/Strategies from Evaluators
Barriers to Evaluation	• Time • Energy to organize • Financial investment • Human resources	• Use existing program data and methods that complement the PGME environment
	• Program is incorrectly implemented • Lack authority to make changes to their program	• Evaluate program implementation fidelity prior to evaluating program impact • Focus stakeholder questions on meaningful program areas over which they hold authority to make changes
	• Findings are unclear	• Interpret the findings alongside the stakeholders and co-develop a manageable follow-up plan
	• Findings remain unused	• Encourage programs to share findings, which establishes an accountability cycle to make changes based on findings
	• Unclear policy expectations	• Advocate for clearer program evaluation policy expectations from the RCPSC
Facilitators to Evaluation	• Engaged stakeholders	• Collaborate as internal evaluation experts with stakeholders who will be using the findings at the beginning of program design and implementation
	• Appropriate evaluation approach	• Lead an evaluation approach that is manageable (e.g., participatory evaluation approach, complement the CBD implementation context)
	• Methodology that leads to publication • Shared resources	• Use evaluation processes that lead to the publication of findings • Offer professional development sessions (e.g., evaluation resources and expertise)

Results

Of the 52 evaluators that we contacted, 12 consented to an interview. Below, we provide a summary of the interview findings where "E" represents an evaluator followed by a number between 1–12 indicating their interview order to preserve anonymity. The themes for the first research question are listed in order of predominance in the data. Whereas the themes for the second and third research questions are listed in a logical order to facilitate how evaluators can help PGME program stakeholders build their capacities to conduct program evaluations and use evaluation findings. Table 1 summarizes barriers and facilitators to evaluation, as well as evaluators' solutions and strategies.

How are evaluators supporting PGME programs with program evaluation?

In the sections that follow, we describe how the participants are currently or could be supporting PGME programs with program evaluation. Specifically, evaluators are or could be: (a) responding in a reactive way as temporary and external evaluation consultants; (b) collaborating with stakeholders as internal evaluation experts from an affiliated university; (c) leading an evaluation approach that is manageable for program directors with limited time; (d) using methods that complement the PGME environment; (e) using evaluation processes that lead to the publication of findings; and (f) offering professional development sessions about program evaluation to support accreditation activities.

Responding in a reactive way as temporary and external evaluation consultants

Participants indicated that they are currently supporting PGME programs in program evaluation as an afterthought, that is, as temporary or external evaluation consultants. Evaluators explained that in their “consultant role, it’s much more reactive because it’s really based on ‘I’ve got a program I want you to help me with,’ whether it be research or loosely... evaluation” [E5]. One problem with “being a reactive, external evaluator who is invited in after program design and implementation is that the evaluation can be seen as a formality rather than a use of the findings” [E2]. Evaluators noted that one of the reasons there are temporary and external evaluators is that the Royal College of Physicians and Surgeons of Canada (RCPSC), the national body that establishes standards for the training, certification, and ongoing professional development of medical specialists in Canada, “lacks clear evaluation policies” [E5].

Collaborating with interested stakeholders as internal evaluation experts from an affiliated university

Evaluators indicated that they are currently or could be supporting PGME programs in program evaluation by collaborating with stakeholders (e.g., program director, program administrator, residents, faculty, academic advisors, competence committee members, clinical supervisors) through an affiliated university’s internal evaluation team [E3, E4, E6, E8, E10]. For example, evaluators’ responsibilities are to collaboratively “develop an evaluation matrix, come up with a plan, evaluate their programs or even just individual initiatives within their programs and to work with them over the course of the process” [E2]. Evaluators highlighted limited resources as a challenge because “they also serve the undergraduate level, postgraduate education, and health professions education” [E3]. As PGME is decentralized, the specialty programs are at different stages of CBD implementation,

which “complicates how senior leadership direct their internal evaluation unit” [E3]. Additionally, despite the RCPSC supporting program evaluation, it depends on initiative from already “overloaded” program directors [E1, E2, E4]. Thus, the value of collaborating with stakeholders within an internal unit is limited by resource constraints between undergraduate and PGME programs and by a reliance on overloaded program directors to seek program evaluation support.

Leading an evaluation approach that is manageable for program directors with limited time

Evaluators indicated that they are currently, or could be, supporting PGME programs in program evaluation by leading an evaluation approach appropriate for program directors with limited time. Evaluators led “rapid-cycle evaluation” [E8], a “continual improvement process or developmental evaluation” [E7], or a “realist evaluation of the implementation of CBD” [E10]. Evaluators noted that the “time commitment” required is a significant challenge in program evaluation, but this obstacle is reduced when working with a professional evaluator [E1, E3, E8, E4, E5]. To make program evaluation manageable, the evaluators suggested requesting reasonable stakeholder buy-in without overloading them (e.g., collaborating on the evaluation matrix, then primarily using email communication until the technical report was under review) and prompting for “short-term and intermediate outcomes” [E5, E7, E8].

Using methods that complement the PGME environment

Evaluators indicated that they are currently, or could be, supporting PGME programs in program evaluation by leveraging methods suited to the overloaded PGME environment without requiring additional resources. Using a standardized process for qualitative and quantitative data collection, they stated that stakeholders’ experiences “informed the design of the interview guide for interviews and focus groups” [E4, E8, E10] and “for the surveys” [E4, E5, E10]. However, participants indicated that programs are “asking for trends in data, and this was going to be quantitative, but the data was not enough to perform a statistical analysis ... in an accurate way” yet [E4]. Successful program evaluation data collection requires a “dedicated personnel position or group who meet weekly to gather and analyze data” [E8]. Faculty burnout was a challenge because the “implementation of CBD has left program directors with minimal time to conduct program evaluations using the same resources and budget” [E2, E4, E6, E7]. Thus, evaluators question the sustainability of evaluation work because it is “often provided by two people using grant funding” [E4]. However, “when program evaluation was considered a priority by the administration, funding was secured” [E5, E8]. In

summary, the evaluators discussed that program evaluation in the PGME context requires methods that complement the busy environment without additional resources.

Using evaluation processes that lead to the publication of findings

The evaluators indicated that they are currently, or could be, supporting PGME programs in program evaluation by assisting program directors in applying evaluation processes that lead to the publication of findings. For example, the evaluators stated that the evaluation processes used in “the technical reports feed into the publication [because] they’ve got the methods nicely standardized” [E8]. They emphasized that their team followed a standardized process for writing a report, sharing findings, and replicating a similar evaluation across other program specialties [E8]. The evaluator’s role is to help develop an “evaluation matrix”, “implement the evaluation process through reporting, and share the “structured evaluation findings with other programs by publishing them in academic journals” [E2, E8, E10]. Publication was one reason program directors were enticed to conduct program evaluation activities and use the findings.

Offering professional development sessions about program evaluation to support accreditation activities

Evaluators indicated that they are currently, or could be, supporting PGME programs in program evaluation by providing professional development to support accreditation. Evaluators often teach evaluation at “postsecondary locations” and “facilitate professional development workshops for various health professions and work settings” [E1, E2]. Furthermore, at the Royal College program evaluation summit, presentations by stakeholders from various universities discussed program evaluation in medical education for interested program directors. Interest in program evaluation is largely driven by the need to meet “accreditation requirements” [E1, E6, E7]. Furthermore, “between accreditation cycles, internal site visits and program surveys contribute to program evaluation” [E3]. In summary, the evaluators emphasized the importance of providing professional development sessions on program evaluation to support accreditation.

How can evaluators help PGME programs build their capacities to do program evaluation?

In the sections that follow, we describe how evaluators can help PGME programs build their capacity to conduct program evaluation. Specifically, evaluators can: (a) use a participatory evaluation approach; (b) offer evaluation resources and expertise as guidance; (c) leverage existing data; and (d) advocate for clearer program evaluation policy expectations from the RCPSC.

Use a participatory evaluation approach

The evaluators indicated that they could help PGME programs build their capacities to do program evaluation by adopting a participatory evaluation approach. With a participatory approach, programs “learn about evaluation by doing it” [E3]. Engaging stakeholders in the “process to learn by doing offers an opportunity for evaluation capacity building” [E5]. A participatory approach builds the capacity of programs to do their own evaluations because “it is impossible for a two-person evaluation team to support 60 plus programs” [E4]. When using a participatory approach, evaluators emphasized breaking tasks into “manageable”, “specific”, and “achievable chunks” [E3, E5]. When “planning with the end in mind”, they indicated that they encourage program directors to focus on overall objectives and work backwards using evaluative language [E5]. Some evaluators leveraged this learning opportunity to “co-create and scaffold an evaluation matrix”, generating “appropriate and accessible program evaluation questions” [E4, E5].

Offer evaluation resources and expertise as guidance

The evaluators indicated that they could help PGME programs build their capacity for program evaluation by providing evaluation resources and expertise to maximize program directors’ limited time. Program evaluation of CBD revealed that “there were competing demands for attention for program directors that influence their ability to effectively implement CBD” [E10]. Thus, the evaluators suggested that it is unlikely programs that do not receive resource support from their PGME office will engage in CBD *evaluation* if they are struggling with CBD *implementation*. Evaluators emphasized the provision of “simplified evaluation resource templates”, “frameworks”, “common approach”, “methodology or tools” (e.g., questions, data collection tools) as a roadmap towards establishing a culture of evaluation and to save time [E1, E3, E6]. However, even with a common framework for program evaluation within PGME programs, evaluators emphasized that resource availability requires a credentialed evaluator with technical evaluation expertise, as well as time and training in data analysis and reporting. Working alongside a credentialed evaluator can help stakeholders ask the appropriate questions, gather data to inform their decision-making, and avoid wasting time.

Leverage existing data

Evaluators indicated that they could help PGME programs build their capacities for program evaluation by maximizing existing data to facilitate the evaluation processes. It is “not necessarily collect[ing] more data, but we build in the time to stop and look at the data that they already have from a systematic basis and think about it differently” [E3]. The RCPSC emphasized the potential

to “exploit existing assessment data” (e.g., “Entrustable Professional Activities”, “the One45 medical education” software platform used to manage clinical rotations, curriculum, and assessments and evaluations, exams and In-Training Evaluation Reports [ITERs]) in program evaluation to improve learning and to address program evaluation questions [E3, E4, E7]. However, “improved technology infrastructure is required to access existing assessment data for program evaluation” [E4]. Thus, the evaluators discussed how they can help PGME programs build their capacity for program evaluation by leveraging existing data with an adequate technology platform to simplify the process.

Advocate for clearer program evaluation policy expectations from the RCPSC

Evaluators indicated that they could help PGME programs build their capacity to conduct program evaluation by raising awareness of program evaluation’s power through set policies. Evaluators advocated for “explicit expectations from the RCPSC and from PGME departments” [E9]. Additional communication is required from the “Canadian Evaluation Society (CES) to inform PGME departments of the evaluation profession to meet stakeholders’ needs and demonstrate fiscal responsibility” [E1]. Moreover, evaluators insisted that deans and program directors require more information about how to leverage program evaluation experts. Clearer policy expectations would result in “better communication between evaluation units and program specialties/sub-specialties” and “distinguish program evaluation from assessment”, as well as quality control (i.e., the process of identifying problems and addressing them) and accreditation [E4, E10]. In summary, evaluators noted program evaluation conversations will occur when the RCPSC, PGME programs, and program directors understand its value as documented in policy.

How can evaluators help PGME program stakeholders build their capacities to use evaluation findings?

In the sections that follow, we describe how the evaluators can help PGME programs build their capacities to use evaluation findings. Specifically, evaluators can: (a) recommend that program evaluators be included at the beginning of program design and implementation; (b) include the stakeholders who will be using the findings in the evaluation itself; (c) focus stakeholder questions on meaningful program areas over which they hold authority to make changes; (d) encourage the use of program evaluation approaches that are appropriate to the CBD implementation context; (e) interpret the findings alongside the stakeholders and co-develop a manageable follow-up plan; and (f) encourage programs to share

findings, which establishes an accountability cycle to make changes based on findings.

Recommend that program evaluators be included at the beginning of program design and implementation

Evaluators indicated that they could help PGME programs build their capacities to use evaluation findings by recommending that program evaluators be incorporated at the beginning of program design and implementation. Evaluators noted that “the best model of evaluation is when you have the evaluator at the planning table... [and] involved up front rather than after the fact” [E6] to listen to and be involved in the stakeholder conversations. However, they noted that one challenge in evaluating CBD implementation is “a program director’s is already overloaded” with the workload that accompanies CBD implementation [E6]. Nevertheless, evaluators insisted that implementation rollout could be smoother with appropriate evaluative planning. Conducting program evaluations as an afterthought often leads to limited use of evaluation findings, as it is “unlikely that additional financial resources would be allocated to undo recent efforts” [E2]. Thus, evaluators recommend that program evaluators be included at the beginning of program design and implementation.

Include the stakeholders who will be using the findings in the evaluation itself

Evaluators indicated that they could help PGME programs build their capacity to use evaluation findings by including the stakeholders who will later use the findings during the evaluation process. Evaluators insisted that stakeholders need to have some level of involvement in the evaluation (e.g., shared decision-making, updates, and engagement throughout the evaluation). Clarifying the degree of stakeholder participation establishes buy-in and understanding and increases the likelihood that the “findings will be used for program improvement” [E2]. However, some evaluators acknowledged that some program directors may prefer to implement a general programmatic approach before asking about CBD specifically. Nevertheless, evaluators emphasized the inclusion of appropriate stakeholders who will utilize the findings in the evaluation process.

Focus stakeholder questions on meaningful program areas over which they hold authority to make changes

The evaluators indicated that they could help PGME programs build their capacity to use evaluation findings by focusing stakeholder questions on areas that stakeholders care about and over which they hold the power to make changes. Evaluators explained that asking evaluation questions that stakeholders care about at the beginning of the evaluation generated stakeholder buy-in.

They added that “if there’s no buy-in for the evaluation to occur from the get-go. Well, then it’s reasonable to consider that then why would they then actually use the evaluation findings?” [E2]. Furthermore, evaluators indicated that questions should investigate areas over which stakeholders hold authority to make changes, and these areas will differ at various hierarchical levels (i.e., the RCPSC, PGME, and program director). Specifically, at the RCPSC level, evaluators suggested using “purposeful evaluation questions to document the combination of the large and small” [E6]. Working closely with the “PGME department and its chair is important to address the questions that these leaders want to have answered” [E5]. At the program director level, evaluators can start with “foundational questions on the implementation delivery domain” [E6] rather than with “system design problems” [E10]. Therefore, evaluators can focus stakeholder questions on meaningful program areas over which stakeholders hold authority to make changes.

Encourage the use of program evaluation approaches that are appropriate to the CBD implementation context

The evaluators indicated that they could help PGME programs build their capacity to use evaluation findings by applying program evaluation approaches appropriate to the CBD implementation context. For example, evaluators explained that using “a realist evaluation recognizes that problems with implementation are not the responsibility of an individual” [E10]. They noted that when evaluation focuses more on learning and improvement than on criticism, program directors are more likely to use the findings. In addition, they stated that other program evaluation approaches that encourage the use of findings are practical participatory and collaborative evaluation approaches because leadership feels invested in the process, which yields the use of findings. Finally, to encourage the use of findings, some participants noted that they applied logic models and “the Kirkpatrick Model to explore the reactions, learning, behaviour, and results” [E4]. However, evaluators asserted that an appropriate CBD evaluation approach should evaluate implementation fidelity prior to evaluating outcomes because the theory should not be accountable for the outcome when the program is implemented incorrectly. Thus, building capacity to use evaluation findings involves approaches that are appropriate to the CBD implementation context.

Interpret the findings alongside the stakeholders and co-develop a manageable follow-up plan

The evaluators indicated that they could help PGME programs build their capacities to use evaluation findings by interpreting the findings with the stakeholders and co-creating a manageable follow-up plan. They insisted that evaluators should participate in the summary and

interpretation of quantitative and qualitative data collection and analysis to ensure that stakeholders “understand findings which increases their likeliness to use the findings” [E1]. Evaluators indicated that stakeholders who use findings well can “share their successes with other programs (e.g., train the trainer)”, creating an evaluation culture that uses the findings [E5]. Participants also emphasized the importance of evaluators “co-developing a formalized follow-up plan” during their committee meetings, one that is manageable for stakeholders [E2]. This approach to institutionalizing a co-created evaluation cycle serves as a follow-up reminder for the next evaluation round, ensuring that challenges are addressed.

Encourage programs to share findings, which establishes an accountability cycle to make changes based on findings

Evaluators indicated that they could help PGME programs build their capacity to use evaluation findings by encouraging programs to share findings, which develops an accountability cycle to inform adjustments based on the findings. Evaluators identified sharing findings through “regular check-ins”, “electronic communication”, or “RCPSC program evaluation summits” as ways to keep stakeholders accountable for applying the findings, noting that there may be future check-ins [E1, E2, E7]. Additionally, they explained that sharing areas of implementation successes and lessons learned from rigorous program evaluation builds “a stronger PGME community” and establishes a more effective Canadian PGME system that learns from its programs [E10]. Moreover, evaluators noted that “involved patients, families, community organizations, and governments develop an evaluation mindset that speaks to social accountability mandates” [E6]. However, depending on the research approach (i.e., a realist evaluation), the findings may not be “integrated into an intentional programmatic cycle because they were instead used to contribute to knowledge about how to conduct program evaluation in medical education” [E10]. Evaluators noted that sharing findings about research on program evaluation also established accountability among evaluators to use these findings the next time they support a specialty/subspecialty program evaluation.

Discussion

In this study, we explored how evaluators support PGME programs and contribute to building evaluation capacity during Canada’s transition to CBD. Although this shift demands substantial investments of time, energy, financial, and human resources from the RCPSC, program directors, and faculty, there remains limited evidence demonstrating whether CBD effectively produces better-prepared physicians or improves patient outcomes compared to traditional models [20, 46–48]. Our findings

reinforce the essential role of evaluators for understanding and guiding this complex transformation.

Evaluators play a critical role in helping PGME programs overcome barriers to evaluation, leverage evaluation facilitators, understand the CBD transition, and provide deeper insights into both implementation processes and the contextual factors that influence outcomes. The evaluators in this study emphasized the necessity of exploring program fidelity before evaluating outcomes, recognizing that program theory (e.g., CBD) cannot be held accountable for unachieved results if implementation is flawed. Evaluators also highlighted how evaluative rigour can, for example, be integrated within PGME programs through common evaluation frameworks, structured evaluations, and standardized methodologies.

The literature identifies three common approaches to program evaluation in medical education that address distinct stakeholder needs: objectives-oriented, process-oriented, and participant-oriented [49]. Objectives-oriented approaches, such as contribution analysis and the Kirkpatrick Model, focus on predefined instructional goals and determining causal relationships [50–52]. While frequently used, evaluators in our study noted that these are most effective after confirming program implementation fidelity, thus aligning with the need first to establish whether PGME programs are being enacted as intended before evaluating their impact on, for example, resident preparedness and patient care.

Process-oriented approaches, including the Context, Input, Process, Product (CIPP) model, rapid cycle evaluation, and realist evaluation, apply systems thinking to monitor development and emphasize ongoing program improvement over proving specific outcomes [48, 49, 52, 53]. Our findings suggest that the transition to CBD naturally lends itself to process-oriented approaches, with evaluators leveraging the integration of rapid cycle evaluations into PGME programs to reduce workload for program directors and generate actionable insights. Some also used realist evaluation, grounded in complexity theory, to unpack the contextual “what works, when, and why” in diverse PGME settings.

Participant-oriented approaches, such as practical participatory evaluation, focus on stakeholders’ perspectives and foster ongoing cycles of data collection and interpretation to build evaluation capacity [49, 54–58]. Evaluators emphasized the value of including program directors early in the evaluation process to enhance buy-in and ensure relevance, given the demanding environment of medical education. However, since busy stakeholders, such as residents and faculty, may lack the time or skills to independently interpret evaluation findings, professional evaluators’ guidance is crucial for transforming data into meaningful program improvements [19, 41]. Despite the benefits of participant-oriented approaches

(e.g., learning through participation in evaluation activities, enhanced stakeholders’ confidence and competence, increased use of findings [29]), evaluators must respect the client who pays for their work [59, 60]. As evaluators are obligated to conduct the kind of evaluation requested by the client, client approval may be required prior to an evaluator applying the solutions and strategies from Table 1. To encourage stakeholder participation, evaluators should be proactively engaged in developing evaluation plans during program development.

Notably, evaluators advocated for consistent terminology and the need to distinguish “program evaluation” from “quality control” or “assessment” and “accreditation,” which can create confusion among stakeholders [61]. Clarification around these terms can help with ECB. ECB is a foundational theme in the literature, but remains underexplored in medical education [13, 28, 29, 31, 42, 62–64]. Our study also revealed that evaluators can encourage evaluative thinking by offering practical resources, including templates, frameworks, and data collection tools and by helping programs use existing data sources. However, they acknowledged the resource limitation as a challenge to ECB.

Organizational and systemic factors significantly influence the sustainability and effectiveness of program evaluation in PGME programs. Evaluators working in internal evaluation units reported precarious job security and unsustainable workloads, as well as the impossibility of supporting overall PGME programs. This instability reflects a broader lack of prioritization and stable funding for evaluation. Clear policies and governance structures are needed at the RCPSC, PGME, and program director levels to support consistent evaluation practices. Professional associations, such as the Canadian Evaluation Society, also play a vital role in advocating for evaluation resources and enhancing communication about the benefits of evaluation. Consistent with best practices in evaluation, our findings reinforce that evaluations must include intended users in decision-making, employ appropriate methods, and focus on questions that are meaningful and actionable within stakeholders’ spheres of authority [34, 35, 65].

Limitations

In total, 12 of the 52 evaluators (23.1% response rate) that we contacted agreed to participate in an interview. This response rate may be viewed as a limitation; however, our interview response rate from evaluators aligns with the survey response rates reported in the literature for evaluators [66, 67]. Another limitation is the potential for evaluators to self-exclude (i.e., self-selection bias) from the interview due to the email subject line, “Program Evaluation in CBME [competency-based medical education] Interview Request,” in our initial request [68]. After

following up with these individuals, some still declined the interview, but their experience in evaluating medical education more generally could have contributed to more informed findings. Finally, evaluators may have provided socially desirable responses due to our presence in the interviews or may be more active in their professions as evaluators and interested in the topic area of program evaluation in medical education and thus, do not accurately represent evaluators in general, resulting in selection bias [68]. Nevertheless, these interview responses provided insight into how evaluators currently support or could support PGME programs in program evaluation, as well as how evaluators can help PGME programs build their capacities for program evaluation and utilize evaluation findings.

Conclusions

Through interviews with medical education evaluators, this study documented how evaluators currently support medical education programs in program evaluation. It also demonstrates how evaluators can assist medical education programs in developing their capacity for program evaluation and utilizing evaluation findings. This documentation presents empirical research on program evaluation in medical education, as well as information for evaluators embarking on evaluations of medical education programs. While this study confirmed that some programs leverage qualified evaluators, the limited internal evaluation units within Canadian medical education universities are concerning. Moreover, within these few internal evaluation units, it is unrealistic to have a small number of evaluators supporting all PGME specialties/subspecialties. If the potential of evaluation is to be realized and sustainable in medical education, allocating adequate resources is essential. Considering fiscally challenging times, such resources and supports can leverage technology to share evaluation templates that model common questions across specialties/subspecialties. Professional development sessions on program evaluation can use technology to make the RCPSC Program Evaluation Summit accessible. Additionally, uniting the existing Canadian medical education internal units can offer an organized and systematic evaluation focus for assessing specialty/subspecialty programs (i.e., either through a divide-and-conquer method or focusing on one specialty/subspecialty each year). Finally, selecting the appropriate evaluation type, questions, and approaches can support program directors and evaluators in continuing to transform medical education.

Abbreviations

AEA	American Evaluation Association
CBD	Competence by Design
CBME	Competency-based medical education
CES	Canadian Evaluation Society

CIPP	Context, Input, Process, Product
ECB	Evaluation capacity building
PGME	Postgraduate medical education
RCPSC	Royal College of Physician and Surgeons of Canada
TIG	Targeted interest group

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12909-026-09239-y>.

Supplementary Material 1.

Acknowledgements

Not applicable

Authors' contributions

JM and KM analyzed and interpreted the evaluator interview data. JM and KM were equal contributors in writing the manuscript. JM and KM read and approved the final manuscript.

Authors' information

We write as researchers on program evaluation in PGME. JM is a part-time professor at a Faculty of Education in Canada, and KM is a full professor in the same faculty in Canada.

Funding

The authors have no sources of funding for the research to report.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

As this study involved human participants, the researchers received ethics approval from the university and consent from evaluator participants. This study was submitted to and approved by an institutional review board (IRB). The full name of the ethics committee is Social Sciences and Humanities Research Ethics Board at the University of Ottawa, Canada. The reference number is S-01–21-6537. The consent that was obtained from all of the participants was informed. This study adhered to the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 25 August 2025 / Accepted: 14 April 2026

Published online: 20 April 2026

References

1. Royal College of Physicians and Surgeons of Canada. CBD implementation. 2025. <https://www.royalcollege.ca/en/standards-and-accreditation/competence-by-design/implementation>. Accessed 3 Jun 2025.
2. Wong L, Sacoransky E, Hopman W, Islam O, Chung AD, Kwan BYM. Radiologist preferences for faculty development initiatives to improve resident feedback in the era of competency-based medical education. *Med Educ Online*. 2024;29:1–12. <https://doi.org/10.1080/10872981.2024.2357412>.
3. Royal College of Physicians and Surgeons of Canada. Competence by Design. 2026. <https://www.royalcollege.ca/en/standards-and-accreditation/competence-by-design.html>. Accessed 2 Mar 2026.
4. Hall AK, Oswald A, Frank JR, Dalseg T, Cheung WJ, Cooke L, et al. Evaluating competence by design as a large system change initiative: readiness, fidelity,

- and outcomes. *Perspect Med Educ*. 2024;13:95–107. <https://doi.org/10.5334/pme.962>.
5. Royal College of Physicians and Surgeons of Canada. CBD program evaluation. 2025. <https://www.royalcollege.ca/en/standards-and-accreditation/competence-by-design/program-evaluation.html>. Accessed 3 Jun 2025.
 6. Van Melle E, Gruppen L, Holmboe ES, Flynn L, Oandasan I, Frank JR. Using contribution analysis to evaluate competency-based medical education programs: it's all about rigor in thinking. *Acad Med*. 2017;92(6):752–8. <https://doi.org/10.1097/ACM.0000000000001479>.
 7. Moreau KA, Cousins JB. A survey of program evaluation practices in family-centered pediatric rehabilitation settings. *Eval Program Plann*. 2014;43:1–8. <https://doi.org/10.1016/j.evalprogplan.2013.10.002>.
 8. Rich JV, Fostaty Young S, Donnelly C, Hall AK, Dagnone JD, Weersink K, et al. Competency-based education calls for programmatic assessment: but what does this look like in practice? *J Eval Clin Pract*. 2020;26:1087–95. <https://doi.org/10.1111/jep.13328>.
 9. Yang A, Newhook D, Sutherland S, Moreau K, Eady K, Barrowman N, et al. Including patients and caregivers in assessment in the pediatric competence by design curriculum: a national consensus study. *Med Teach*. 2023;45:604–9. <https://doi.org/10.1080/0142159X.2022.2152661>.
 10. Crawford L, Cofie N, McEwen L, Dagnone D, Taylor SW. Perceptions and barriers to competency-based education in Canadian postgraduate medical education. *J Eval Clin Pract*. 2020;26:1124–31. <https://doi.org/10.1111/jep.13371>.
 11. Moreau KA. Exploring the connections between programmatic assessment and program evaluation within competency-based medical education programs. *Med Teach*. 2021;43:250–2. <https://doi.org/10.1080/0142159X.2020.1841128>.
 12. Royal College of Physicians and Surgeons of Canada. CBME program evaluation summit. 2024. <https://www.royalcollege.ca/ca/en/cbd/understanding-cbd/cbd-program-evaluation/cbme-program-evaluation-summit.html>. Accessed 15 Aug 2024.
 13. Milosek J, Eady K, Moreau KA. Program evaluation in competence by design: a mixed-methods study. *J Med Educ Curric Dev*. 2025;12. <https://doi.org/10.1177/23821205251321791>.
 14. Harachi TW, Abbott RD, Catalano RF, Haggerty KP, Fleming CB. Opening the black box: Using process evaluation measures to assess implementation and theory building. *Am J Community Psychol*. 1999;27:711–31. <https://doi.org/10.1023/A:1022194005511>.
 15. Milosek J. Evaluation in competence by design medical education programs. University of Ottawa. 2023. <https://ruor.uottawa.ca/server/api/core/bitstream/s/ab20cf92-7c3a-4783-8162-261ffa9c37ba/content>.
 16. Vassar M, Wheeler DL, Davison M, Franklin J. Program evaluation in medical education: an overview of the utilization-focused approach. *J Educ Eval Health Prof*. 2010;7(1):1–3. <https://doi.org/10.1033/jeehp.2010.7.1>.
 17. American Evaluation Association. AEA evaluator competencies. 2025. <https://www.eval.org/About/Competencies-Standards/AEA-Evaluator-Competencies>. Accessed 10 Aug 2025.
 18. Canadian Evaluation Society. Evaluator competencies. 2024. <https://evaluationcanada.ca/career/evaluator-competencies.html>. Accessed 4 Jun 2024.
 19. Gruppen LD, Ten Cate O, Lingard LA, Teunissen PW, Kogan JR. Enhanced requirements for assessment in a competency-based, time-variable medical education system. *Acad Med*. 2018;93:517–21. <https://doi.org/10.1097/ACM.0000000000002066>.
 20. Canadian Evaluation Society. Evaluation standards. 2025. <https://evaluationcanada.ca/career/evaluation-standards.html>. Accessed 2 Jun 2025.
 21. Fetterman DM. Empowerment evaluation and social justice: confronting the culture of silence. New York: Guilford Press; 2023.
 22. Nkwake AM. Constituents of evaluation practice. In: Credibility, validity, and assumptions in program evaluation methodology. 2nd edition. Switzerland: Springer; 2023. p. 1–24.
 23. Patton MQ. Facilitating evaluation to enhance use. In: Facilitating evaluation: principles in practice. Thousand Oaks: Sage Publications; 2018. p. 27–42.
 24. Alkin MC, Vo AT, Christie CA. Evaluation essentials: from A to Z. 3rd edition. New York: Guilford Press; 2024.
 25. Stake RE. Standards-based evaluation. In: Standards-based & responsive evaluation. Thousand Oaks: Sage Publications; 2004. p. 54–85.
 26. Yarbrough DB, Shulha LM, Hopson RK, Caruthers FA. The program evaluation standards: a guide for evaluators and evaluation users. 3rd ed. Thousand Oaks: Sage Publications; 2011.
 27. Naccarella L, Pirkis J, Kohn F, Morley B, Burgess P, Blashki G. Building evaluation capacity: definitional and practical implications from an Australian case study. *Eval Program Plann*. 2007;30:231–6. <https://doi.org/10.1016/j.evalprogplan.2007.05.001>.
 28. Parsons BL, Lovato CY, Hutchinson K, Wilson D. Building evaluation capacity through CLIPs: communities of learning, inquiry, and practice. *New Dir Eval*. 2016;2016:21–36. <https://doi.org/10.1002/ev.20199>.
 29. Preskill H. Evaluation's second act: a spotlight on learning. *Am J Eval*. 2008;29:127–38. <https://doi.org/10.1177/1098214008316896>.
 30. Stockdill SH, Baizerman M, Compton DW. Toward a definition of the ECB process: a conversation with the ECB literature. *New Dir Eval*. 2002;2002:7–26. <https://doi.org/10.1002/ev.39>.
 31. Cousins JB, Bourgeois I. Multiple case study methods and findings. *New Dir Eval*. 2014;2014:25–99. <https://doi.org/10.1002/ev.20077>.
 32. Bourgeois I, Cousins JB. Understanding dimensions of organizational evaluation capacity. *Am J Eval*. 2013;34:299–319.
 33. Patton MQ, Campbell-Patton C. Utilization-Focused Evaluation. 5th ed. Thousand Oaks: Sage Publications; 2021.
 34. Cousins JB, Earl LM. The case for participatory evaluation: theory, research, practice. In: Participatory evaluation in education: studies of evaluation use and organizational learning. Washington: Falmer Press; 1995. p. 3–20.
 35. Hotte N, Simmons L, Beaton K, LDCP Workgroup. Scoping review of evaluation capacity building strategies. Cornwall: Public Health Ontario; 2015.
 36. Preskill H, Boyle S. Insights into evaluation capacity building: motivations, strategies, outcomes, and lessons learned. *Can J Program Eval*. 2008;23:147–74. <https://doi.org/10.3138/cjpe.0023.008>.
 37. Creswell J, Creswell J. Research design: Qualitative, quantitative, and mixed methods approaches. 6th edition. Thousand Oaks: Sage Publications; 2023.
 38. Patton MQ. Qualitative evaluation and research methods. 2nd ed. Newbury Park, CA: Sage Publications; 1990.
 39. Patton MQ. Qualitative research & evaluation methods: integrating theory and practice. 4th edition. Thousand Oaks: Sage Publications; 2015.
 40. Holmboe ES, Ward DS, Reznick RK, Katsfrakis PJ, Leslie KM, Patel VL, et al. Faculty development in assessment: the missing link in competency-based medical education. *Acad Med*. 2011;86:460–7. <https://doi.org/10.1097/ACM.0b013e31820cb2a7>.
 41. Oandasan I, Martin L, McGuire M, Zorzi R. Twelve tips for improvement-oriented evaluation of competency-based medical education. *Med Teach*. 2020;42:272–7. <https://doi.org/10.1080/0142159X.2018.1552783>.
 42. Cousins JB, Goh SC, Clark S, Lee LE. Integrating evaluative inquiry into the organizational culture: a review and synthesis of the knowledge base. *Can J Program Eval*. 2004;19:99–141. <https://doi.org/10.3138/cjpe.19.006>.
 43. King JA. Developing evaluation capacity through process use. *New Dir Eval*. 2007;2007:45–59. <https://doi.org/10.1002/ev.242>.
 44. Seidman I. Interviewing as qualitative research: a guide for researchers in education and the social sciences. 5th ed. New York: Teachers College Press; 2019.
 45. Miles MB, Huberman AM, Saldaña J. Qualitative data analysis: a methods sourcebook. 4th ed. Thousand Oaks: Sage Publications; 2020.
 46. American Evaluation Association. What is evaluation? 2025. <https://www.eval.org/About/What-is-Evaluation>. Accessed 6 Jun 2025.
 47. Fitzpatrick JL, Sanders JR, Worthen BR, Wingate LA. Introduction to program evaluation: alternative approaches and practical guidelines. 5th edition. Upper Saddle River: Pearson; 2023.
 48. Frank JR, Karpinski J, Sherbino J, Snell LS, Atkinson A, Oswald A, et al. Competence by design: a transformational national model of time-variable competency-based postgraduate medical education. *Perspect Med Educ*. 2024;13:201–23. <https://doi.org/10.5334/pme.1096>.
 49. Cook D. Twelve tips for evaluating educational programs. *Med Teach*. 2010;32:296–301. <https://doi.org/10.3109/01421590903480121>.
 50. Kirkpatrick Partners. The Kirkpatrick Model. 2024. <https://www.kirkpatrickpartners.com/the-kirkpatrick-model/>. Accessed 11 Aug 2025.
 51. Moreau KA, Eady K. Connecting medical education to patient outcomes: the promise of contribution analysis. *Med Teach*. 2015;37(11):1060–2. <https://doi.org/10.3109/0142159X.2015.1060307>.
 52. Stufflebeam DL, Coryn CLS. Evaluation theory, models, & applications. 2nd edition. San Francisco: Jossey-Bass; 2014.
 53. Hall AK, Rich J, Dagnone JD, Weersink K, Caudle J, Sherbino J, et al. It's a marathon, not a sprint: rapid evaluation of competency-based medical education program implementation. *Acad Med*. 2020;95(5):786–93. <https://doi.org/10.1097/ACM.0000000000003040>.
 54. Cousins JB. Will the real empowerment evaluation please stand up?: a critical friend perspective. In: Fetterman DM, Wandersman A, editors. Empowerment evaluation principles in practice. New York: Guilford Press; 2005. p. 183–208.

55. Cousins JB, Earl LM. The case for participatory evaluation. *Educ Eval Policy Anal.* 1992;14:397–418. <https://doi.org/10.3102/01623737014004397>.
56. Cousins JB, Whitmore E. Framing participatory evaluation. *New Dir Eval.* 1998;1998:5–23. <https://doi.org/10.1002/ev.1114>.
57. Moreau KA. Twelve tips for planning and conducting a participatory evaluation. *Med Teach.* 2017;39:334–40. <https://doi.org/10.1080/0142159X.2017.1286310>.
58. Patton MQ. A utilization-focused approach to contribution analysis. *Evaluation.* 2012;18:364–77. <https://doi.org/10.1177/1356389012449523>.
59. Scriven M. Evaluation as a discipline. *Stud Educ Eval.* 1994;20:147–66. [https://doi.org/10.1016/S0191-491X\(00\)80010-3](https://doi.org/10.1016/S0191-491X(00)80010-3).
60. Scriven M. Evaluating evaluations: A meta-evaluation checklist. *Consultado em.* 2011;9:1–7.
61. CanRAC. General standards of accreditation for residency programs. Ottawa: CanRAC; 2020.
62. Labin SN, Duffy JL, Meyers DC, Wandersman A, Lesesne CA. A research synthesis of the evaluation capacity building literature. *Am J Eval.* 2012;33:307–38. <https://doi.org/10.1177/1098214011434608>.
63. Sarti AJ, Sutherland S, Landriault A, DesRosier K, Brien S, Cardinal P. Understanding of evaluation capacity building in practice: a case study of a national medical education organization. *Adv Med Educ Pract.* 2017;2017:761–7. <https://doi.org/10.2147/AMEP.S141886>.
64. Steinert Y. Faculty development in the health professions: a focus on research and practice. 2nd edition. Singapore: Springer; 2025.
65. Patton MQ. The sociological roots of utilization-focused evaluation. *Am Sociol.* 2015;46:457–62. <https://doi.org/10.1007/s12108-015-9275-8>.
66. Wanzer DL. Research on evaluation: it takes a village (the problem). 2018. <http://danawanzer.com/roe-theproblem/>. Accessed 24 Aug 2022.
67. Wanzer DL. What is evaluation?: perspectives of how evaluation differs (or not) from research. *Am J Eval.* 2020;42(1):28–46. <https://doi.org/10.1177/1098214020920710>.
68. Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. Boston: Houghton Mifflin; 1979.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.