# Evaluation in Competence by Design Medical Education Programs

Jenna D. Milosek

Thesis submitted to Dr. Katherine Moreau, Dr. Ruth Kane, Dr. Peter Milley, and Dr. Timothy Wood in partial fulfillment of the requirements for the degree of Doctor of Philosophy of Education

Faculty of Education
University of Ottawa

**Abstract**

To ensure medical residents are prepared to work in complex and evolving settings, postgraduate medical education is transitioning to competency-based medical education, which is known as Competence by Design (CBD) in Canada. To understand how CBD is operationalized within specific residency programs and how it contributes to patient, faculty, and learner outcomes, there is a need to engage in program evaluation. However, the actual extent that, reasons for, and methods in which CBD programs are engaging in program evaluation remain unclear. Furthermore, minimal attention has been given to building program evaluation capacity within medical education programs (i.e., doing evaluation and using evaluation findings).

In this research project, I explore and formally document: (a) the extent that and the ways in which CBD programs are engaging in program evaluation, (b) the reasons why these programs are engaging or not engaging in program evaluation, (c) the actual and potential positive and negative consequences of these programs engaging in program evaluation, (d) the ways that these programs build their capacities to do program evaluation and use evaluation findings, (e) the ways that program evaluators currently support these programs, and (f) the ways that program evaluators can help stakeholders build their capacities to do program evaluation and use evaluation findings. Through this research, I contribute to the limited body of empirical research on program evaluation in medical education. Confirming how CBD programs are engaging in program evaluation can advise stakeholders and program evaluators on how best to support CBD programs in building their capacities to do program evaluation and use evaluation findings, inform the design and implementation of other medical education programs, and, ultimately, enlighten program evaluation research on authentic and current evaluation practices in medical education.

To meet the objectives of this study, I used a three-phase, sequential mixed methods approach. In Phase 1, I conducted a survey of Canadian program directors whose programs have transitioned to CBD to determine: (a) the extent to which CBD programs are engaging in program evaluation, and (b) the reasons why CBD programs are engaging or not engaging in program evaluation. In Phase 2, I interviewed interested program directors to explore: (c) how CBD programs are engaging in program evaluation, and (d) the ways in which CBD programs can build their capacities to do program evaluation and use evaluation findings. In Phase 3, I

interviewed Canadian program evaluators to investigate: (e) how program evaluators are currently supporting CBD programs in program evaluation, and (f) how program evaluators can help CBD programs build their capacities to do program evaluation and use evaluation findings.

Overall, the Phase 1 findings show that: (a) over three quarters of respondents indicated that their program does engage in program evaluation and most invite stakeholders to participate. However, most programs rarely leverage the expertise of a program evaluator and acknowledge interpreting quantitative program evaluation data is a challenge. Additionally, (b) most programs engage in program evaluation to improve their program and make decisions. However, most programs do not have an employee whose primary responsibility is program evaluation. They do not receive funding for program evaluation which affects their abilities to engage in program evaluation. Moreover, some programs do not engage in program evaluation because they do not know how to do program evaluation. The Phase 2 findings show that: (c) when program directors do engage in program evaluation, they are using ad hoc evaluation methods and a team-based format. However, program directors of CBD programs are struggling to engage in program evaluation because of limited available resources (i.e., time, financial, human resources, and technology infrastructure) and buy-in. Additionally, (d) program directors are building their capacity to do evaluation and use the findings from their specialty/subspecialty program evaluation. The Phase 3 findings show that: (e) program evaluators are supporting CBD programs by responding in a reactive way as temporary and external evaluation consultants. Finally, (f) program evaluators can help CBD programs build their capacities to do program evaluation and use the findings by using a participatory evaluation approach, leveraging existing data, encouraging the use of program evaluation approaches that are appropriate to the CBD implementation context, or encouraging programs to share findings which establishes an accountability cycle. In light of these findings, I discuss ways to engage in program evaluation, build capacity to do evaluation, and build capacity to use evaluation findings in CBD programs.
*Keywords*: program evaluation, medical education, competence by design, evaluation capacity building

**Acknowledgements**

I am forever grateful for the support, direction, and patience I received from my supervisor, Dr. Katherine Moreau. Your remarkable work ethic and constant reinforcement has encouraged me and has impacted my life journey. I appreciate your assistance during this challenging endeavor more than I can express. I would also like to thank Dr. Ruth Kane, Dr. Peter Milley, and Dr. Timothy Wood, who have helped to develop my academic mindset and who have shaped this thesis. In addition, I am grateful for how Dr. Ruth Kane, Dr. Kaylee Eady, and Dr. Angus McMurtry have filled my graduate studies with opportunities to learn and grow as a research assistant and teaching assistant. Also, thank you Dr. Eady for reviewing my survey and interview guides. Thank you to Dr. Sarah Heath for reviewing my interview guides as well. I am thankful for the opportunity to engage in graduate studies, which has been made possible by the generous scholarships from the University of Ottawa Admission and Excellence Scholarship and the Ontario Graduate Scholarship. Thank you for this learning journey.

I would also like to thank my friends and family for their encouragement throughout this process. Thank you to my mother and father, Patricia and Dennis Milosek, who reassured me throughout my graduate studies. Thank you to my mother-in-law and father-in-law, Ursula Irzykowska and Jan Ozimierski, for giving me strength to persevere. Thank you to my sister, Jackie Milosek, for reminding me of my determination. Finally, thank you to my husband, Dominic Ozimierski, for your support during my ongoing studies and for believing in me. I dedicate this thesis to all of you.

## Table of Contents

## List of Appendices

## List of Tables



## List of Figures

## List of Abbreviations

| Abbreviation | Explanation |
|---|---|
| AEA | American Evaluation Association |
| CBD | Competence by Design |
| CBME | Competency-based medical education |
| CIPP | Context, Input, Process, Product |
| ECB | Evaluation capacity building |
| EPA | Entrustable professional activities |
| MMR | Mixed methods research |
| PE | Participatory Evaluation |
| PGME | Postgraduate medical education |
| RCPSC | Royal College of Physician and Surgeons of Canada |
| TIG | Targeted interest group |

# Chapter 1: Introduction

## Statement of Problem

To ensure that medical residents are prepared to work in complex and evolving settings, postgraduate medical education is transitioning to competency-based medical education (CBME; Royal College of Physicians and Surgeons of Canada [RCPSC], 2020b). CBME is an approach "to preparing physicians for practice that is fundamentally oriented to graduate outcome abilities and organized around competencies derived from an analysis of societal and patient needs" (Frank et al., 2010). In Canada, medical education uses the term Competence by Design (CBD) to describe how CBME is (or will be) applied in 67 specialty and subspecialty Canadian postgraduate training programs (RCPSC, 2020b). The objective of CBD is to train physicians who are competent to meet local health demands and to improve patient care by enhancing learning and assessment in residency training (RCPSC, 2022). The intent of CBD is not to change the duration of residency, but to use learning time more effectively. Under this approach, postgraduate medical education programs must demonstrate that a resident is competent in essential areas of clinical practice before moving to the next training stage. While this approach does not dictate the teaching and assessment strategies that programs must use, it does define a set of abilities or milestones that residents must attain (Iobst et al., 2010). It is in contrast to traditional medical education that is subject-centered and uses time-based curricula (Shah et al., 2016). In the Canadian context, CBD is characterized as a hybrid education model that integrates a time-based and an outcomes-based approach to learning (RCPSC, 2022). The RCPSC explains that CBD "reviews the design, implementation, assessment and evaluation of each specialty program across Canada's 17 medical universities, using CanMEDS 2015 as an organizing framework of competencies" (RCPSC, 2022). However, despite this transition to CBD, there is minimal evidence to support whether CBD develops increasingly prepared physicians or improves patient outcomes when compared to traditional medical education (Van Melle, Gruppen, et al., 2017).

This lack of evidence stems, in part, from a lack of systematic program evaluations of CBD programs. Program evaluation is "the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming" (Patton, 1997, p. 23). In order to understand how CBD is operationalized within specific residency

programs and how it contributes to patient, faculty, and learner outcomes, there is a need to engage in program evaluation. Evaluations of CBD, when methodically planned and implemented, can determine the effectiveness of CBD programs and ensure that residents receive high-quality education. Moreover, the medical education literature demonstrates that program evaluation is essential because it provides vital information about: (a) the types of educational programs needed, (b) whether programs are being implemented as intended, (c) how programs can be improved, and (d) whether certain programs should continue (Cook, 2010). However, research has shown that program evaluation in medical education often lacks proper forethought, structure, and methodology (Vassar et al., 2010). The Core Components Framework (Van Melle et al., 2019) offers a common structure to guide evaluation on whether CBD is being implemented as intended. The intentional linking of theory and practice in the CBD Core Components Framework provides a significant step towards a deeper understanding of the conditions in which CBD most effectively enhances patient care outcomes (Van Melle et al., 2019). Despite this useful common framework, the *actual* extent that, the reasons why, and the methods in which CBD programs are engaging in (or not engaging in) program evaluation remain unclear. Furthermore, minimal attention has been given to building program evaluation capacity within medical education programs, such as augmented motivation, knowledge and skills to improve programs, and organizations' abilities to do evaluation and to use evaluation findings (Labin et al., 2012).

There are numerous opportunities to do evaluation and use program evaluation findings for CBD programs. Program evaluations can occur at the level of the RCPSC, postgraduate medical education (PGME) departments, or within a specific specialty program. To guide a program evaluation, participating stakeholders (e.g., program director, faculty, program administration, residents, and clinical site-supervisors) can develop major evaluation questions centering on, for example, the need for the program, program theory and design, processes, outcomes, efficiency, or impact (Rossi et al., 2019). Stakeholders then have the power to reflect on and make changes to the programs based on the evaluation findings.

The literature reviewed for this study revealed a need for further systematic program evaluations of CBD programs. It also provided examples of the types of evaluation approaches that programs may use. For example, authors have shown that rapid evaluations provide insights into the successes and challenges of CBD implementation (Hall et al., 2020). Moreau and Eady

(2015) also highlighted how contribution analysis may be a useful evaluation approach, as the program is viewed as part of a causal package that works together with other factors, interventions, and influences to bring about changes in observed outcomes. Moreover, participatory evaluation (PE), which involves collaboration between trained program evaluators and stakeholders, can generate meaningful findings and recommendations to improve medical education programming and provide CBD stakeholders with access to real-time evaluation data that can inform immediate program changes (Moreau, 2017).

To understand and improve program evaluation within CBD, it is important to investigate the extent to which it is occurring as well as strategies for building evaluation capacity. Specifically, research must investigate and document: (a) the extent that and the ways in which CBD programs are engaging in program evaluation, (b) the reasons why these programs are engaging or not engaging in program evaluation, (c) the actual and potential positive and negative consequences of these programs engaging in program evaluation, (d) the ways that these programs build their capacities to do program evaluation and use evaluation findings, (e) the ways that program evaluators currently support these programs, and (f) the ways that program evaluators can help stakeholders build their capacities to do program evaluation and use evaluation findings. Such an investigation can help to improve program evaluation in CBD and clarify the relationships (or lack thereof) between educational objectives and outcomes, explore the use of evaluation findings to inform program decision making, and determine programs' values or worth (Mertens & Wilson, 2012). It can also increase the theoretical understanding of how program evaluation works within CBD (Van Melle, Frank, et al., 2017), thereby facilitating the delivery of quality medical education that benefits residents, faculty, health professionals, and patients. By documenting program evaluation practices within CBD, this study also builds on the empirical research on program evaluation and program evaluation capacity building within CBD. Furthermore, it will provide recommendations for the development of sustainable program evaluation practices in CBD and will influence those who are developing and delivering CBD to incorporate effective program evaluation practices into their programs.

Through my three-phase mixed methods thesis, I will build on the limited body of empirical research on program evaluation in CBD. I will expand our understanding of the use of program evaluation to evaluate CBD programs and provide important information to CBD decision-makers and those who evaluate medical education programs. The findings of this

project will inform the RCPSC of the success and challenges of embedding program evaluation within residencies and will support program directors in deciding how to approach systematic and iterative evaluation of their CBD programs.

**Contributions to the Field**

This study contributes to supporting program evaluation in medical education. It is imperative to investigate how program evaluation is being applied within medical education, especially within CBD. The goal is to learn about the actual extent that, the reasons why, and the methods in which CBD programs are engaging in (or not) program evaluation. There may be a gap between program evaluation scholarship that informs medical education and how medical education is in fact applying program evaluation. Therefore, confirming how CBD programs are engaging in program evaluation can advise stakeholders and program evaluators on how to best support CBD programs in building their capacities to do program evaluation and use evaluation findings, inform the design and implementation of other medical education programs, and, ultimately, enlighten program evaluation research on authentic and current evaluation practices in medical education.

**Overview of Thesis**

This thesis includes eight chapters. Chapter 1 introduces the problem, gaps in the literature, and requirements for further research in this area. Chapter 2 offers a detailed literature review to explore the ideas and conceptual framework of this study and specifies the research questions. The literature review incorporates empirical research on the purposes of program evaluation within CBD, applications of certain recommended program evaluation approaches, the roles of program evaluators in the evaluation of CBD programs, and capacity building to do program evaluation and to use the findings. Chapter 3 explains the methodological design, my position (i.e., background as an elementary school teacher engaging in professional development which led to my interest in program evaluation in a new context), the philosophical assumptions embedded in this study, and a description of the mixed methods used in the three-phase research design. Chapters 4 through 6 highlight the study findings. Chapter 4 describes the results of Phase 1, which surveyed program directors of CBD programs who are involved in the implementation of their CBD program. Chapter 5 explains the findings of Phase 2, which leveraged one-on-one interviews with program directors to explore their use of (or lack thereof) program evaluation in their CBD program. Chapter 6 documents Phase 3, which included one-

on-one interviews with program evaluators within PGME departments. It also included program evaluators who were members of the American Evaluation Association (AEA) Topical Interest Groups (TIGs) for Health Evaluation or Health Professions Education, Evaluation and Research, and who either evaluate CBD programs or do not evaluate CBD programs but may offer insight into applicable strengths and limitations. Chapter 7 integrates and discusses the key findings from all three phases of the study and connects these findings to relevant published empirical research about program evaluation in medical education. Chapter 8 identifies the study's limitations and areas of possible future research and presents concluding statements on the use of program evaluation in CBD.

# Chapter 2: Literature Review and Conceptual Framework

In this chapter, I review two bodies of literature with two specific purposes. First, I explore the current conversations in the peer-reviewed literature on program evaluation in medical education and define the key terms for this study. To justify the need for investigating the topic of program evaluation in medical education, I synthesize the purposes of program evaluation within CBD, applications of certain recommended program evaluation approaches, and the roles of program evaluators in the evaluation of CBD programs. Lastly, I examine the research literature on evaluation capacity building (ECB) to assess the nature and extent of current knowledge in this area. Since there is minimal literature on ECB within medical education, this section reviews ECB as well as the concept of evaluation use, which is essential for ECB. I conclude this chapter by presenting the specific research questions that I investigated during this study and the conceptual framework that I developed from the literature presented in this chapter and within which I contextualized my analyses.

## Program Evaluation in Medical Education

The literature on program evaluation in medical education and CBD suggested several purposes of program evaluations, recommended the application of certain program evaluation approaches, and encouraged medical educators to seek support from trained program evaluators in the evaluation of CBD programs.

### *Purposes of Program Evaluations in Medical Education*

Stakeholders are interested in doing program evaluation in medical education for a variety of purposes. For example, they may conduct program evaluations for accountability or learning purposes (Cousins et al., 2015). In terms of accountability, stakeholders can assess program effectiveness, identify best practices, and validate their programs within financially competitive climates (Clarke, 2006). Alkin and Christie (2012) noted that there are three types of accountabilities in program evaluation, including goal accountability, process accountability, and outcome accountability. Each type refers to a specific aspect of the program and is the responsibility of a particular stakeholder group. Goal accountability focuses specifically on whether relevant program goals have been developed and is of particular interest to governing boards and upper-level management (e.g., RCPSC). Process accountability examines the implementation and appropriateness of the procedures used to obtain program goals and is the responsibility of faculty and learners. Conversely, outcome accountability explores the degree to

which goals have been achieved and is also the responsibility of faculty and learners (Alkin & Christie, 2012).

Program evaluation for learning purposes within medical education allows all stakeholders to grow and develop together. As Torres et al. (2005) noted, through participating actively in program evaluation, individuals can increase their competencies in data collection, analysis, and communication (behaviourism). They can also begin to understand issues more completely and gain greater appreciation of issues (constructivism) or increase their abilities to think differently about program problems or issues through the application of new cognitive clues (cognitive learning; Torres et al., 2005). Stakeholders within medical education can come together as a team to collaboratively explore program issues or problems, thus obtaining a more diverse and complex understanding of their programs. Furthermore, they are able to learn about each other's perspectives and encourage learning (e.g., across specialties and subspecialties that have transitioned to CBD).

The literature revealed that few researchers have investigated the purposes of program evaluation in medical education. Nevertheless, those who have published in this area stated that, in terms of accountability, medical educators and learners were able to document short-, medium-, and long-term outcomes attributable to programs (Van Melle, Frank, et al., 2017). Program evaluation for accountability purposes can also provide stakeholders with insight into the "black box" of program functioning (Funnell & Rogers, 2011). That is, it can clarify whether learners' outcomes are attributable to program theory, objectives, and/or implementation. Moreover, program evaluation researchers within medical education noted that program evaluation provides educators with knowledge about their programs and ongoing foci for ongoing program development (Goldie, 2006). Within medical education, there is a need for rich, practical, and detailed understandings about which educational interventions work, for whom, in what circumstances, and why (Van Melle, Frank, et al., 2017). In 2019, the RCPSC completed two program evaluation studies on the implementation of the 2017 and 2018 CBD launch disciplines and the level of readiness of the 2019 CBD launch disciplines (RCPSC, 2019). These two studies offered residents, faculties, and supervisors opportunities to learn about CBD implementation from each other and identify challenges, including the pace of culture change affecting aspects of CBD (e.g., giving and receiving feedback), Entrustable Professional Activities (EPAs), milestones application, workload increase, the impact of CBD on stress and

well-being, training and information, and electronic platform creation (RCPSC, 2019). Other medical education disciplines commencing their CBD journeys were able to learn from these evaluation findings and use them in the planning and implementation of their own programs.

**Common Program Evaluation Approaches in Medical Education**

As medical educators consider how they are engaging in program evaluation, it is essential to recognize stakeholders' evaluation needs, to contemplate the intended uses of the eventual evaluation findings, and to adopt appropriate evaluation approaches (Cook, 2010). Some program evaluation researchers within medical education have recommended specific approaches for evaluating medical education programs, namely objectives-oriented, process-oriented, and participant-oriented approaches (Cook, 2010). In objectives-oriented approaches, the instructional goals are defined at the beginning of the activity and evaluated at the end to determine if the specified goals have been met (Cook, 2010). Here, reductionism assumes a cause–effect linear relationship in the program's elements that can be understood and predicted by investigating the contribution of the constituent parts (Stufflebeam & Shinkfield, 2007). Process-oriented approaches apply general systems theory to specify if a programming need exists, outline the most appropriate way to satisfy that need, and track the development process during implementation (Cook, 2010). Finally, participant-oriented approaches address how the stakeholders involved in a program view it, use an ongoing cycle of data collection and interpretation, and apply triangulation from multiple perspectives (Cook, 2010). When focusing on participants, complexity theory has been used to contextualize the reciprocal relationships between program stakeholders within their environments and how those environments may influence the participants (Doll & Trueit, 2010).

Contribution analysis exemplifies an objectives-oriented approach that is theory-based for evaluating the performance of programs, and it is ideal for the evaluation of medical education programs. It focuses on evaluating the contribution that a program is making to observed outcomes (e.g., learner outcomes, patient outcomes). Rather than showing causation between a program and observed outcomes, it aims to reduce uncertainty regarding the difference a program has made on specific outcomes (Mayne, 2011). Contribution analysis uses six steps, including:

- setting out the cause-and-effect issue(s) that need exploration through evaluation,
- developing a proposed theory of change,

- gathering existing evidence to support the theory of change,
- assembling a story of contribution,
- searching for additional evidence to support the theory of change and story of contribution, and
- revising and strengthening the theory of change and story of contribution (Mayne, 2019).

As Moreau and Eady (2015) noted, contribution analysis views a medical education program as part of a causal package that works together with other factors, interventions, and influences to bring about changes in outcomes. The main evaluation question when using contribution analysis focuses on whether the medical education program is a necessary part of the causal package to bring about the given change in the observed outcomes.

The New World Kirkpatrick Model is another objectives-oriented evaluation approach that includes four levels (i.e., reaction, learning, behaviour, and results). The first level considers participants' reaction, satisfaction, and engagement with the program and its relevance to their jobs and responsibilities (Kirkpatrick Partners, 2020). The second level emphasizes learning and participants' acquisition of the intended knowledge, abilities, attitudes, confidence, and dedication based on their training (Kirkpatrick Partners, 2020). The third level focuses on behaviour, the application of the program training to participants' career responsibilities, and the required drivers as systems that reinforce performance of critical behaviours on the job (Kirkpatrick Partners, 2020). Finally, level four highlights results as the degree to which targeted outcomes occur based on training and leading indicators reveal short-term measurements to suggest that behaviours are heading in a positive direction (Kirkpatrick Partners, 2020). The New World Kirkpatrick Model maintains the four levels of the original Kirkpatrick model (Kirkpatrick, 1959; Kirkpatrick, 1976; Kirkpatrick, 1996; Kirkpatrick, 1998; Kirkpatrick & Kirkpatrick, 2006) but it combats criticisms of the original model and addresses the complexities of current learning environments.

Alternatively, the Context, Input, Process, Product (CIPP) model is a process-oriented approach to program evaluation that focuses on program improvement instead of attempting to prove an aspect about the program (Stufflebeam & Shinkfield, 2007). It does not assume linear constraining direction, but evaluates an educational program based on complex and dynamic relationships (Stufflebeam & Shinkfield, 2007). The CIPP approach includes four

complementary sets of evaluation studies (e.g., context, input, process, and product evaluation) that permit evaluators to review essential but often overlooked program dimensions (Frye & Hemmer, 2012). Together, the CIPP model components are flexible to the ever-changing nature of educational programming and can provide program improvement data (Stufflebeam & Shinkfield, 2007). By alternately focusing on a program's context, inputs, process, and products, the CIPP model addresses each educational programming phrase including planning, implementation, improvement-focused, and summative reflection (Stufflebeam & Shinkfield, 2007). A context evaluation study is useful when planning a new program, adapting an established program to changing circumstances (e.g., new leadership), or making decisions about removing existing programs (Stufflebeam & Shinkfield, 2007). An input evaluation is applicable for resource allocation (e.g., staff, budget, time) in planning an educational program, assessment of feasibility or cost-effectiveness to changes, program development to respond to unfolding program needs, and justification for funding and resources to a new program (Stufflebeam & Shinkfield, 2007). A process evaluation study is used to assess program implementation and to prepare the evaluator to interpret the outcomes by highlighting how the elements impact outcomes (Stufflebeam & Shinkfield, 2007). The attention to process issues encourages ongoing data collection to inform program management and ongoing effective change within the complexity of medical education programs (Frye & Hemmer, 2012). Product evaluation focuses on identifying the positive/negative, intended/unintended, and short-term/long-term program outcomes as well as their impact and sustainability, and the transportability of the program (Frye & Hemmer, 2012).

Finally, PE is a participant-oriented approach, which is commonly advocated for in medical education. It includes two streams: practical participatory evaluation (P-PE) and transformative participatory evaluation (T-PE). The primary objective of P-PE is "to foster program decision making and problem solving and the use of evaluation findings and processes" (Cousins, 2005, p. 185). It is geared towards formative program evaluations, which improve stakeholders' understanding and implementation of their programs (Cousins & Earl, 1992). Usually, P-PE only involves primary stakeholders who are in leadership roles within programs and can thus use the evaluation findings to facilitate programming changes. Conversely, the main purpose of T-PE is to empower community groups that are oppressed by dominating groups (Cousins & Whitmore, 1998). It typically involves program beneficiaries in evaluation processes,

is stakeholder-driven, and enables the evaluator to take on a purely facilitative role. Regardless of the stream selected, PE involves active stakeholder participation (e.g., residents, administrators, faculty, clinician educators) in program evaluation. This participation enriches a sense of ownership over the evaluation, embeds evaluative thinking into an organization's culture, and enhances the utilization of evaluation findings for those who will use them (Patton, 2012). Thus, within medical education, PE can improve the utilization of findings, empower medical organizations and individuals to take charge of evaluation activities, and build necessary evaluation capacity within organizations and among individuals who must maintain continuing evaluation activities (Moreau, 2017). Based on its collaborative emphasis between trained program evaluators and stakeholders, PE can also generate meaningful findings and recommendations to improve medical education programming (Moreau, 2017). Moreover, it is useful within CBD, as it can provide stakeholders with access to real-time evaluation data that informs their immediate program changes (Moreau, 2017).

### *Support from Program Evaluators to Evaluate CBD Programs*

Experienced program evaluators can help with the evaluation of CBD programs as they are experts with experience in reflective, technical, situational, management, and interpersonal practice skills that are necessary when planning, investigating, and applying evaluation findings (Canadian Evaluation Society, 2018). It is unrealistic to expect CBD residents and faculty to recognize and apply constructive program evaluation feedback without support or skill development (Gruppen et al., 2018). Successful evaluations need to be navigated through targeted planning and evaluation alongside an experienced evaluator that highlights practice, clarifies feedback, and encourages persistence among faculty and residents (Holmboe et al., 2011). Oandasan et al. (2020) suggested enlisting a professional evaluator (i.e., through the Canadian Evaluation Society, AEA, or International Organization for Cooperation in Evaluation) to support the planning and implementation of a CBD program evaluation. Regardless of an evaluator's familiarity with CBD, an outside perspective is often necessary to see how a comprehensive evaluation can inform CBD (Oandasan et al., 2020).

## ECB in Medical Education

In contrast to an imposed, one-time evaluation, ECB applies evaluation theory, practice, and policy to facilitate an ongoing and sustainable learning community (Preskill, 2008). Additionally, ECB is a tool that can strengthen and sustain future program evaluation activities

by developing an organization's ability to design and implement evaluations as well as stakeholders' abilities to gain and use their evaluation knowledge and skills (King, 2007). The contextualized, intentional system of guided practices can bring about continuous and evolving quality program evaluation for use within one or more organizations and programs (Stockdill et al., 2002). The focus of ECB is to equip staff at the individual and organizational levels with the knowledge and skills to complete rigorous evaluations and ensure that such evaluations become embedded as routine practice (Naccarella et al., 2007). Individuals have their own ways of contributing to ECB by participating in the evaluation and making use of the findings. When encountering ever-changing program funding, ECB is a strategy to review services and improve on program effectiveness (Compton et al., 2008). Specifically, evaluative thinking becomes an ongoing organizational learning experience to identify an area for improvement, gather data, generate meaning, and shape practice (Parsons et al., 2016).

There are two types of ECB: (a) direct ECB involving planned ECB activities within and outside of evaluation projects, such as training on data analysis, and (b) indirect ECB that is the product of stakeholder involvement in evaluation processes that lead to building evaluation knowledge (Cousins et al., 2004). Direct ECB is an intentional process to augment individual motivation, knowledge, and abilities while enhancing a group's competencies towards conducting or using evaluation (Labin et al., 2012). Indirect ECB is the learning function of program evaluation that focuses on building stakeholders' understanding of program evaluation and processes, instead of focusing solely on program evaluation findings. It is beneficial because it changes individuals' technical abilities (i.e., to make sense of findings) and perspectives of evaluation. ECB in medical education includes building capacities to do program evaluation and to use program evaluation findings.

**Capacities to Do Program Evaluation**

The capacity to do program evaluation refers to the knowledge and skills needed to carry out evaluation activities. Cousins et al. (2008) clarified that these knowledge and skills "pertain to evaluation planning, standards of professional evaluation practice, instrument development, data collection, analysis and interpretation, and reporting and follow-up" (p. 5). Furthermore, capacities to do evaluation are reflected in human resources, organizational resources, and evaluation planning and activities. Human resources include the staffing of evaluation positions within an organization (e.g., medical school, residency programs), the evaluation unit's

leadership quality, and evaluators' technical and interpersonal skills, such as identifying evaluation issues and using appropriate data collection tools, their internal and external professional development activities, and their ongoing skill assessments (Bourgeois & Cousins, 2013). Organizational resources are based on the evaluation budget's stability, ongoing data collection for performance measurement systems, and organizational infrastructure that supports evaluation policies and governance structure (Bourgeois & Cousins, 2013). Evaluation planning and activities involve external supports from professional associations and the distribution of evaluation results to members of the unit (Bourgeois & Cousins, 2013).

***Strategies***

Sometimes applied independently or in overlapping ways, there are different teaching and learning strategies that relate to the organization, its systems, its staff, and its evaluation approach that can help individuals learn about and build their capacities to do program evaluation (Cousins & Bourgeois, 2014). Involvement in a program evaluation process (e.g., advisory committee, collaborative evaluation approaches) develops experiential learning from the planning and implementation of program evaluation (Preskill & Boyle, 2008). By participating in program evaluations, individuals can gain access to ECB resources (i.e., staff and time), exposure to data collection and technology monitoring systems, and connections to evaluation templates, which they can apply towards maintaining ongoing and future evaluations (Hotte et al., 2015). Stakeholders' access to formal training (e.g., workshops, exposure to high-level evaluation theorists and practitioners) or informal training (e.g., through involvement in evaluation processes) builds confidence to practice evaluation (Cousins & Bourgeois, 2014). Furthermore, technical assistance from an internally or externally contracted experienced evaluator can help individuals who are conducting their own evaluations. Experienced evaluators can share written resources and materials, exposing stakeholders to current evaluation research and scaffolding their inquiry of evaluation (Preskill & Boyle, 2008). An external evaluator offers technical assistance, sustainable partnerships, and an objective perspective to view program relationships in a different way (Hotte et al., 2015). However, instead of contracting technical expertise from outside, an organization can establish a dedicated internal evaluation unit with a trained evaluator to manage evaluation processes, act as a technical expert and facilitator, and lead the planning and analyzing of evaluative data (Cousins & Bourgeois, 2014).

**Capacities to Use Program Evaluation Findings**

Capacities to use evaluation processes and findings encourage evaluation literacy, informed organizational decision making, and learning benefits for stakeholders (Bourgeois & Cousins, 2013). Evaluation literacy refers to involvement in evaluation and results-management orientation because organizations encourage individuals' involvement in the evaluation process (Bourgeois & Cousins, 2013). Informed organizational decision making refers to decisions that are informed by evaluation information, such as findings and suggestions pertaining to budget allocation and policy (Bourgeois & Cousins, 2013). Finally, learning benefits refers to when evaluation discoveries can be the basis for action change, can alter stakeholders' understandings, or can encourage behavioural change (Bourgeois & Cousins, 2013).

The use of evaluation findings can fall under four different categories: (a) conceptual, (b) instrumental, (c) symbolic, and (d) process. Conceptual use is when evaluation findings are used to influence stakeholders' thinking about a program (Alkin & King, 2016). In this sense, it can influence individuals' cognitive processing and how they perceive the program as well as its associated activities. Instrumental use is when program directors use evaluation findings to support decision-making or problem-solving processes (Shulha & Cousins, 1997). This type of use involves applying evaluation findings to support direct actions (e.g., using the findings to make changes in a medical education program). Symbolic use leverages data to comply with reporting demands or to publicly support a decision/opinion made on a different basis (Alkin & King, 2016). Finally, process use is when participation in evaluation enhances use (i.e., a by-product of stakeholder engagement; Harnar & Preskill, 2007). For example, as a result of involvement in the evaluation process, changes can occur in program directors' attitudes, thought processes, and behaviours (Patton, 2007). The effects of process use include facilitating program directors' understandings of evaluation, supporting and reinforcing program intervention, facilitating CBD program development, infusing evaluation thinking into health care culture, and keeping program directors focused on program priorities (Patton, 2007). Building capacity to use evaluation findings encourages accountability for strengthening the relationship between program resources, activities, outputs, and short- and long-term outcomes (Patton, 2012). Greater accountability for program implementation and organizational improvement cycles into further organizational learning and efficiency.

### *Strategies*

Although evaluation findings usually support positive program improvement, sometimes pushback can occur due to various factors (e.g., staff concerns about the termination of their program, fear over scrutinization from upper management, increased time commitment, or no designated position for research and evaluation; Cousins & Bourgeois, 2014). There are strategies that can encourage individuals to build their capacities to use program evaluation findings and improve CBD programs. One strategy to enhance stakeholders in *using* findings complements the *doing* aspect of evaluation. Individuals conducting evaluations must demonstrate adequate technical rigour of methods to justify the findings as legitimate evidence to the users (Cousins & Earl, 1995). Thus, building capacity to *do* evaluation directly impacts the likelihood of building capacity to *use* evaluation findings. Moreover, instead of producing general findings, these findings must be specific to the intended users, who were actively involved and engaged throughout the evaluation. That is, the intended uses of the evaluation findings by the intended users should inform the evaluation decision-making processes (Patton, 2015b). Meaningful results increase the use of findings (e.g., stakeholder involvement, strong stakeholder communication, regular data collection, evaluation designs with comparison groups, and strong participant interest due to minimal pre–post time lapses; Cousins & Bourgeois, 2014). Another strategy for building capacity to use evaluation findings is to establish a cycle of curricula review, development, and implementation, such as annual reviews (Cousins & Earl, 1995). Within a cycle of review, follow-up briefings should be implemented where individuals share their experiences with applying the findings. These briefings are a way to hold individuals accountable as they will be responsible for sharing their successes and challenges of implementing the findings with other stakeholders (Patton, 2008). This follow up offers a platform to model highlights, make connections, and share resources to encourage further use of findings (Patton, 2008). Applying the aforementioned strategies to do program evaluation and to use program evaluation findings contributes to building evaluation capacity in medical education.

To support program evaluation in medical education, further research is required to learn more about the actual extent of and the reasons for CBD programs engaging (or not) in program evaluation. Learning more about how CBD programs are engaging in program evaluation and building their capacities to do program evaluation and use evaluation findings is necessary to

strengthen the relationship between program evaluation scholarship that listens to and informs medical education. Furthermore, confirming how CBD programs are engaging in program evaluation can advise program evaluators on ways to support CBD programs and build their capacities to do program evaluation and use evaluation findings.

**Summary and Research Questions**

In this section, I restate the objectives of this thesis and introduce the research questions that will be explored. In the last subsection, I present and outline the conceptual framework derived from the literature discussed above that I used to steer this thesis. Overall, the literature revealed that although program evaluation may be complicated when used to evaluate CBD programs, it can be useful for evaluating program implementation, effectiveness, and outcomes. For example, as mentioned above, program evaluation of CBD has the potential to: (a) provide stakeholders with information to assess program effectiveness, identify best practices for training physicians, and validate their programs within financially competitive climates; (b) examine the implementation and appropriateness of the procedures used for obtaining program goals; and (c) increase medical education stakeholders' and programs' capacities to do and use evaluation. These potentialities highlight the need to empirically investigate and formally document: (a) the extent that and the ways in which CBD programs are engaging in program evaluation, (b) the reasons why these programs are engaging or not engaging in program evaluation, (c) the actual and potential positive and negative consequences of these programs engaging in program evaluation, (d) the ways that these programs build their capacities to do program evaluation and use evaluation findings, (e) the ways that program evaluators currently support these programs, and (f) the ways that program evaluators can help stakeholders build their capacities to do program evaluation and use evaluation findings. Given these gaps, issues, and opportunities, I focused on learning more about program directors' experiences with program evaluation in CBD and the ways in which program evaluators are supporting CBD programs to build their capacities to do program evaluation and use evaluation findings. Thus, the following six research questions guided my three-phase mixed methods research (MMR) project:

**Phase 1 (Survey of Canadian Program Directors of CBD):**

1. To what extent are CBD programs engaging in program evaluation?
2. Why are CBD programs engaging or not engaging in program evaluation?

**Phase 2 (Interviews with Canadian Program Directors of CBD):**

3. How are CBD programs engaging in program evaluation?
4. How can CBD programs build their capacities to do program evaluation and use evaluation findings?

**Phase 3 (Interviews with Medical Education Program Evaluators):**

5. How are program evaluators currently supporting CBD programs in program evaluation?
6. How can program evaluators help CBD programs build their capacities to do program evaluation and use evaluation findings?

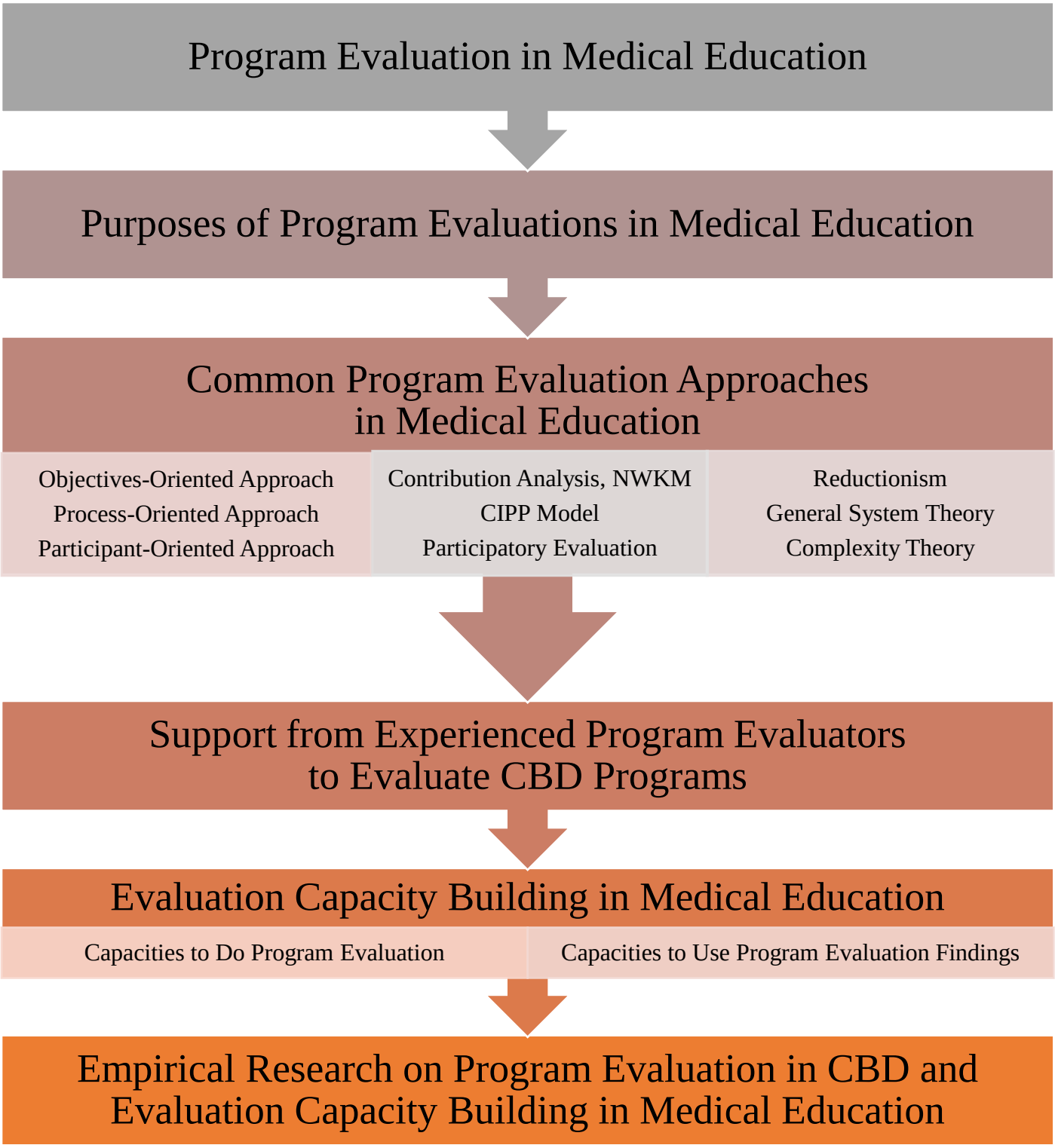
**Conceptual Framework**

In this chapter, I reviewed two bodies of literature: (a) the peer-reviewed literature on program evaluation in medical education, which highlighted the purposes of program evaluation within CBD, the applications of certain recommended program evaluation approaches, and the roles of program evaluators in the evaluation of CBD programs; and (b) the research literature on capacity building to do program evaluation and to use the findings to assess the nature and extent of current knowledge in this area. In the previous subsection, I clarified the research questions that I employed to study the use of program evaluation to evaluate CBD programs. These research questions are based on assumptions derived from the literature, my understanding of that literature, and my position as a program evaluation scholar. These assumptions influenced my study and a diagram of their relationships, and my conceptual framework is provided in Figure 1. As the figure demonstrates, research on program evaluation in medical education serves as the lens of this study. Using this lens, I acknowledge that there are areas of empirical work where program evaluation scholars have conducted research on program evaluation in medical education to contribute to what is known about this area of study. Literature on program evaluation in medical education suggested that the purposes of program evaluations in medical education are: (a) to determine whether change has occurred and to what extent the change can be attributed, (b) to build an evidence base, (c) to clarify the program's value, and (d) to meet accountability demands. Program evaluation scholars and theorists recommend common program evaluation approaches in medical education based on objectives, process, or participants. To facilitate the application of program evaluation approaches, medical education decision-makers are encouraged to seek support from experienced program evaluators to evaluate CBD programs. With trained evaluator support, stakeholders can learn to engage in ECB in medical education. Program evaluators can assist CBD programs in building their

capacities to do program evaluation by directing planning stages, ensuring relationships between the questions posed and evidence gathered, and maintaining momentum throughout the evaluation. The capacity to do evaluation impacts the capacity to use the findings. By focusing on program evaluation in CBD and ECB in medical education, I can create research on program evaluation in CBD to understand the nature and extent of program evaluation in medical education. Ultimately, this work contributes to empirical research on program evaluation in medical education and provides information to medical education decision-makers and program evaluators who are interested in evaluating medical education programs.

**Figure 1**

*Conceptual Framework for Research on Evaluation in CBD*




*Note*. CBD = Competence by Design; NWKM = New World Kirkpatrick Model; CIPP = Context, Input, Process, Product.

# Chapter 3: Methodology

In this chapter, I discuss my epistemology (i.e., what is knowledge) and I synthesize my study's methodology (i.e., how knowledge is constructed; Lincoln & Guba, 2000). I begin by sharing my positionality statement, which clarifies my interest in medical education and how it connects with program evaluation. I then situate program evaluation within medical education under a pragmatic epistemology. I offer a detailed examination of a pragmatist epistemology and how this way of knowing reflects my perspective on knowledge and my relationship with the research context. Next, I examine the design of my three-phase sequential MMR study. Finally, I conclude the chapter with a description of the ethical considerations.

## Researcher Positionality

I am an elementary school teacher. Continuing professional development programs offer teachers opportunities to enhance their teaching abilities and acquire prerequisites to further their educational careers. A common prerequisite for the Principal Qualification Program is a master's degree in education. In January 2017, I completed my master's with a major research paper that focused on research on program evaluation. I pursued my program evaluation research interest using my teaching experience as a platform to gain a deeper understanding of program evaluation in K-12 schools. I then decided to pursue the present doctoral research on program evaluation in medical education, as a way of exploring a new educational context.

## Epistemology: Pragmatism as My Philosophical Approach to Knowledge

A paradigm or worldview is a broad philosophical orientation about the complexity of the real world that a researcher applies to a study (Creswell & Creswell, 2018). A paradigm outlines how a researcher comprehends reality (i.e., ontology), knows what they know (i.e., epistemology), incorporates their values (i.e., axiology), and structures the ways that they determine knowledge (i.e., methodology; Lincoln & Guba, 2000). As a researcher, an epistemology is *what* I perceive as knowledge and *how* I know that knowledge (Davis, 2004). The epistemological approach that I am applying to this study is pragmatism.

Pragmatism is a paradigm that attempts to unite the gap between the scientific and structuralist orientation of older approaches (e.g., postpositivism using quantitative methods) and the naturalistic methods and orientation of newer approaches (e.g., constructivism using qualitative methods; Creswell & Plano Clark, 2018). Pragmatism assumes that "in order to be able to attribute a meaning to concepts, one must be able to apply them to existence" (Dewey,

1998, p. 4). Thus, what is understood as knowledge and the process of how we know what we know is developed through experience. Pragmatism accepts that through interaction with the environment (i.e., forming beliefs, applying actions, and associating consequences), organisms generate knowledge about the world as justified assertions (Dewey, 1998). These assertions are rooted in context, so what constitutes valid and meaningful knowledge can change with time for pragmatists (Johnson & Onwuegbuzie, 2004). That is, a pragmatist's ontology assumes that people have a level of agency, their epistemology applies to solve real-world problems, their axiology values solving real-world problems, and their methodology includes mixed methods—both quantitative and qualitative data collection and analyses.

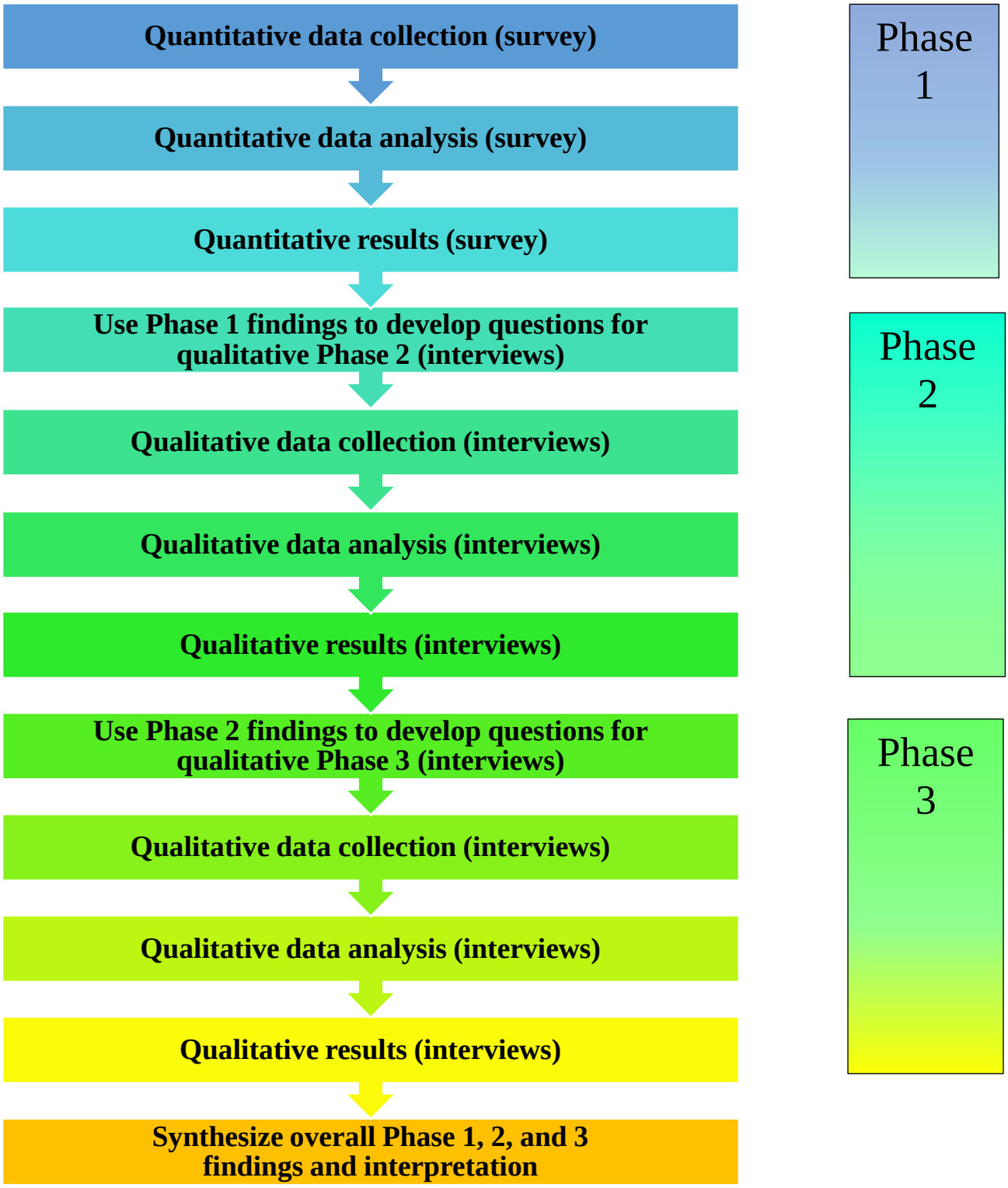
Tashakkori and Teddlie (2003) advocated for using the single paradigm of pragmatism to engage in MMR. Pragmatism highlights that knowledge is "what works" and the practical aftermath of experience and research informing us about the future (Creswell & Plano Clark, 2018). Deciding on what works derives from how objects are (i.e., objectivism) and how the brain perceives them (i.e., subjectivism; Tashakkori & Teddlie, 1998). Pragmatism approves mixing independent paradigms to address certain research problems (Greene & Caracelli, 1997). This mixing is possible as pragmatism supports that a phenomenon can be situated within various realms, such as individuals' inner worlds, behaviour, symbolism, or natural environments (Patton, 1988).

**Methodology: Three-Phase Research Design**

In this study, I followed a three-phase, explanatory sequential research design. Phase 1 focused on the collection and analyses of quantitative survey data, Phase 2 involved the collection and analyses of qualitative interview data, and Phase 3 included qualitative interview data from a different population (Creswell & Creswell, 2018). I collected and analyzed both the quantitative and qualitative data separately; however, I have triangulated these findings in Chapter 7 (Creswell & Plano Clark, 2018). A diagram of the research design is presented in Figure 2.

**Figure 2**

*Research Design Visual Representation – Phases 1, 2, and 3*



### *Defining MMR*

There are many definitions for MMR. For example, Greene et al. (1989) focused on the methods and philosophy aspects in their definition:

> Mixed-method designs are those that include at least one quantitative method (designed to collect numbers) and one qualitative method (designed to collect words), where neither type of method is inherently linked to any particular inquiry paradigm. (p. 256)

In this definition, they refer to MMR as a method and view it is as a "doing tool" for collecting and analyzing data (Giddings & Grant, 2007). Over time, Greene (2006) has further developed her definition of MMR by focusing on the multiple ways that researchers can use it to make sense of the social world. She now defines MMR inquiry as:

> An approach to investigating the social world that ideally involves more than one methodological tradition and thus more than one way of knowing, along with more than one kind of technique for gathering, analyzing, and representing human phenomena, all for the purpose of better understanding. (Greene, 2006, p. 93)

In contrast, other researchers have insisted that MMR belongs within a paradigm and is a "thinking tool" that structures how researchers frame their research questions and select their research methods. Tashakkori and Teddlie (2003) emphasized this perspective by focusing on MMR as a methodology and its evolution to the point where MMR holds a distinct methodological orientation (i.e., applying a separate worldview, vocabulary and techniques). Furthermore, others in the field, such as Creswell and Plano Clark (2018) and myself, have defined MMR as both a method and a philosophy. In this regard, MMR has philosophical assumptions and values that govern the mixing of quantitative and qualitative research approaches to inquiry. It involves the collection, analysis, and integration of quantitative and qualitative data in a single study. Moreover, the word *methods* broadly includes "methods of data collection (e.g., questionnaires, interviews, observations), methods of research (e.g., experiments, ethnography), and related philosophical issues (e.g., ontology, epistemology, axiology)" (Johnson et al., 2007, p. 118). Given this interpretation, the mixing in MMR is complex and involves more than just combining quantitative and qualitative techniques of data collection. With this definition, evaluation researchers can mix quantitative and qualitative data collection techniques to conduct research on how program evaluation is being conducted within CBD contexts.

### *Types of MMR Designs*

There are six types of MMR designs that researchers can use for program evaluation within CBD contexts, including: (a) the convergent parallel design, (b) the explanatory sequential design, (c) the exploratory sequential design, (d) the embedded design, (e) the transformative design, and (f) the multiphase design (Creswell, 2015). Table 1 provides an explanation of each of these designs.

**Table 1**

*Description of Mixed Methods Research (MMR) Designs*

| Type of MMR design | Description |
|---|---|
| Convergent parallel design | Collects quantitative and qualitative data simultaneously. |
| Explanatory sequential design (QUAN + qual) | Starts with quantitative data collection and analysis that is followed up with qualitative data collection and analysis. |
| Exploratory sequential design (QUAL + quan) | Begins with qualitative data collection and analysis that is supported with quantitative data collection and analysis. |
| Embedded design | Includes simultaneously collecting primary data (quan or qual) that is supported by secondary data (quan or qual) either before, during, or after the study. |
| Transformative design | Applies a theoretical framework, such as a feminist, racial, sexual orientation, or disability theory, to guide the quantitative and qualitative study. |
| Multiphase design | Conducts multiple projects over time that build upon one another. |

In this study, I used a three-phase, explanatory sequential research design. Phase 1 involved the collection and analysis of survey data from program directors to examine the extent to which: (a) CBD programs are engaging in program evaluation, and (b) the reasons why CBD programs are engaging or not engaging in program evaluation. Findings from the program director survey informed the development of the Phase 2 qualitative interview guide and protocol that was used to collect data through semi-structured interviews with a selected group of program directors. The purpose of the interviews was to understand the nature and ways in which: (a) CBD programs are engaging in program evaluation, and (b) CBD programs can build their

capacities to do program evaluation and use evaluation findings. The Phase 1 and 2 findings informed the development of the Phase 3 interview guide and protocol for program evaluators in medical education. The goal of the Phase 3 interviews with program evaluators was to determine: (a) the extent to which program evaluators are currently supporting CBD programs in program evaluation, and (b) the ways in which program evaluators can help CBD programs build their capacities to do program evaluation and use evaluation findings. Although Phase 1, 2, and 3 data collection and analysis were completed separately, the findings are triangulated in Chapter 7 (Creswell et al., 2011).

**Ethical Considerations**

I obtained approval for this study from the University of Ottawa Social Sciences and Humanities Research Ethics Board (see Appendix A). All oral and written communication with program directors and AEA member participants detailed the research purpose and clarified confidentiality, anonymity, and the participants' rights and options to withdraw from the study at any time.

**Summary**

In this chapter, I explored an overview of my researcher positionality, the philosophical assumptions, methodology, and ethical considerations present in this study. My reflective positionality as an elementary school teacher encouraged me to pursue professional development programs to enhance my teaching abilities and to further my educational career. After completing my master's degree, which focused on program evaluation in K-12 schools, I pursued this doctoral research on program evaluation in medical education to explore a new educational context. My pragmatic philosophical assumptions shape how I, as a researcher, comprehend reality (i.e., ontology), know what I know (i.e., epistemology), incorporate my values (i.e., axiology), and structure the ways that I determine knowledge (i.e., methodology; Lincoln & Guba, 2000). I am influenced by Dewey (1998) and I apply pragmatism, which accepts that through interaction with the environment (i.e., forming beliefs, applying actions, and associating consequences), I can generate knowledge about research on program evaluation in CBD as justified assertions. I outlined the three-phase explanatory sequential mixed methods design for the study. In Phase 1, I developed a survey to collect mainly quantitative and some qualitative data from Canadian specialty/subspecialty program directors in CBD. The goal was to explore the extent to which CBD programs are engaging in program evaluation and the reasons

why CBD programs are engaging or not engaging in program evaluation. In Phase 2, I conducted semi-structured interviews with a selected group of program directors from Phase 1. Phase 2 focused on how CBD programs are engaging in program evaluation, and how CBD programs are building or can build their capacities to do program evaluation and use evaluation findings. In Phase 3, I conducted semi-structured interviews with program evaluators who have an interest in medical education. Phase 3 focused on how program evaluators are currently supporting CBD programs in program evaluation, and how program evaluators can help CBD programs build their capacities to do program evaluation and use evaluation findings. The next three chapters describe Phase 1, Phase 2, and Phase 3 of my thesis in greater detail, including the sample population, instrument development, data collection and analysis, and the findings from each phase.

# Chapter 4: Phase 1

In this chapter, I describe the methods and results of Phase 1, during which I surveyed program directors from Canadian specialty/subspecialty programs that transitioned to CBD between 2017–2020. The purpose of this phase was to explore the following research questions:

- Question 1: To what extent are CBD programs engaging in program evaluation?
- Question 2: Why are CBD programs engaging or not engaging in program evaluation?

## Sample

All English and French Canadian specialty/subspecialty program directors whose programs had transitioned to CBD received an invitation to participate in the survey. The timing of CBD launch varied across disciplines, as follows:

- 2017: anesthesiology and otolaryngology;
- 2018: emergency medicine, forensic pathology, medical oncology, nephrology, surgical foundations, and urology;
- 2019: critical care medicine, gastroenterology, general internal medicine, geriatric medicine, internal medicine, radiation oncology, rheumatology, cardiac surgery, neurosurgery, obstetrics and gynecology, anatomical pathology, and general pathology; and
- 2020: general surgery, neurology (adult and pediatrics), nuclear medicine, orthopedic surgery, pediatric cardiology, physical medicine and rehabilitation, plastic surgery, and psychiatry (RCPSC, 2020c).

Each Canadian specialty medical program has a designated program director. However, not all Canadian universities offer all of the specialty medical residency programs. Thus, I emailed a total of 445 program directors the survey invitation link to participate. Program directors' email contact information was retrieved from the RCPSC website. The surveys were accessed from a location deemed appropriate by the program director. Three survey completion reminders with the accessible link were sent (i.e., survey sent on May 3, 17, and 31, 2021).

## Instrument Development

I developed the program director survey tool (see Appendix C for the English version and Appendix D for the French) based on the literature review in Chapter 2 and a review of other evaluation scholars' surveys who have investigated evaluation experiences (Cousins et al., 2008).

Two experts (i.e., my supervisor and an evaluation professor) reviewed the program director survey for clarity. I applied changes to improve the wording and administered the survey to a small representative sample (e.g., doctoral peers) to review the obtained results and to gain further insight on wording to ensure clarity and relevance of questions (Lancaster et al., 2004). I created the survey using SurveyMonkey (https://www.surveymonkey.com), which is hosted on a Canadian server, so the survey data are stored in Canada and regulated by Canadian privacy legislation. Table 2 presents a summary of the Phase 1 research questions and associated survey dimensions and questions.

**Table 2**

*Specifications for Phase 1 Survey: Program Directors of CBD Programs*

| Research question | Dimension | Corresponding survey questions |
|---|---|---|
| 1. To what extent are CBD programs engaging in program evaluation? | Specification of whether a CBD program is engaging in program evaluation | 1, 2 |
| | Nature and frequency of CBD programs engaging in program evaluation | 3 |
| 2. Why are CBD programs engaging or not engaging in program evaluation? | Purpose | 4 |
| | Accountability | 5 a, b, c |
| | Learning | 5 d, e, f, g, h |
| | Conceptual use of program evaluation | 6 a, b |
| | Instrumental use of program evaluation | 6 c, d |
| | Symbolic use of program evaluation | 6 e, f |
| | Process use of program evaluation | 6 g, h |
| | Other reasons | 5 i, j, 6 i, 7, 8, 9 |

*Note*. CBD = Competence by Design.

The survey comprised 11 questions, including a consent question to participate in the survey and a question to clarify whether the participant's program engaged in program evaluation to guide the survey direction. Five of the 11 questions were open-ended, offering the participants opportunities to further explain their experiences, while the other six were closed-ended. I included the response selection of "Don't know" to reduce missing responses (Millar & Dillman, 2011).

**Data Collection Procedures**

I sent a personalized, bilingual explanation in an email containing the survey link and information letter (see Appendix E for the English version and Appendix F for the French) to each specialty/subspecialty program director based on the CBD launch schedule (RCPSC, 2020a). In returned surveys, participants' consent to participate in Phase 1 of this study was inferred based on their responses to question 1 (i.e., "I agree" or "I do not agree" to consent to completing this survey). To maximize participation, I emailed the survey using a modified version of Millar and Dillman's (2011) tailored design method to conform to the online context, as detailed in Table 3. The survey took approximately 15 minutes for program directors to complete. To encourage participation and to thank program directors for their participation in the study, each program director received a summary of the findings. If they were interested, participants had the option to enter an email address at the end of the survey to participate in a follow-up interview and receive a $10 Amazon gift card. Survey participants were permitted to skip any questions they did not want to answer, so there was a different number of respondents for each survey question. The total number of Phase 1 survey respondents for each question is noted in Appendix G.

**Table 3**

*Tailored Design Method for Survey Distribution*

| Email | Contact point | Procedure |
|---|---|---|
| 1st email | Day 1 | Initial request: Email to program directors with the Phase 1 study letter of information and survey link for them to complete. |
| 2nd email | Day 14 | First reminder: Email to program directors thanking those who have completed the survey and providing a reminder and link to the survey for those who have yet to complete the survey. |
| 3rd email | Day 28 | Final reminder: Email to program directors thanking those who have completed the survey and providing a reminder and link to the survey for those who have yet complete the survey. |

*Note*. Adapted from "Improving Response to Web and Mixed-Mode Surveys," by M. M. Millar and D. A. Dillman, 2011, *Public Opinion Quarterly*, *75*(2), pp. 249–269. (https://doi.org/10.1093/poq/nfr003). Copyright 2011 by the Author.

## Data Analysis

Using IBM SPSS Statistics (Version 28), I calculated descriptive statistics (e.g., frequencies, percentages) for the 11 survey items. I also completed a content analysis of the five open-ended survey questions that gave respondents the option to share additional insights. First, I read over all the responses for the five open-ended questions and established a list of recurring themes, which I characterized as a string of words with a subject and a predicate. Next, I reviewed the responses again and grouped them according to the identified themes. Finally, I calculated the number of responses and percentages for each within each thematic group. I present the different themes and their frequency in the findings section that follows (Lune & Berg, 2017).

## Findings

### *Consent of Respondents*

A total of 149 program directors consented to complete the survey out of the possible 445 who were invited to participate. Thus, the response rate was 33.5%. Of the 149 respondents, 134 (89.9%) were English program directors and 15 (10.1%) were French program directors.

### *Research Question 1: To What Extent Are CBD Programs Engaging in Program Evaluation?*

Of the 149 respondents, 143 (96.0%) program directors answered the question about whether their CBD program engages in program evaluation, and 6 (4.0%) skipped this question. Over three quarters ($n$ = 127, 88.8%) of respondents indicated that their program does engage in program evaluation, and 16 (11.2%) indicated that their program does not engage in program evaluation. Of the respondents whose programs do not engage in program evaluation, 12 (75.0%) were English program directors and 4 (25.0%) were French program directors. Of the respondents whose programs do engage in program evaluation, 120 (94.5%) were English program directors and 7 (5.5%) were French program directors.

**Survey Question 3a–n: Program Directors' Perceptions About the Frequency of Their CBD Program Engaging in Certain Program Evaluation Activities.** Table 4 presents the perceptions of English and French program directors whose CBD programs engage in program evaluation (question 3a–n). Of the 127 respondents, between 86 and 90 elected to answer various parts of this question about their program's evaluation activities. In total, more than half of the respondents indicated that they either "never" ($n$ = 36/86, 41.9%) or "rarely" ($n$ = 22/86, 25.6%) worked with a program evaluator. Most of the respondents indicated that they

invite stakeholders to participate in program evaluations either “frequently” (*n* = 32/90, 35.6%) or “always” (*n* = 32/90, 35.6%). More than half of the respondents answered that they “rarely” (*n* = 25/88, 28.4%) or “sometimes” (*n* = 24/88, 27.3%) interpret quantitative program evaluation data. Finally, most respondents indicated that they “sometimes” (*n* = 27/87, 31.0%), “frequently” (*n* = 31/87, 35.6%), or “always” (*n* = 24/87, 27.6%) use program evaluation findings.

**Table 4**

*Program Directors' Perceptions About Their CBD Program Engaging in Program Evaluation*

| | Evaluation activity | *n* (%) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Survey respondents . . . | *N* | Never | Rarely | Sometimes | Frequently | Always | Don't know |
| Q3a. | work with a program evaluator. | 86 | 36 (41.9) | 22 (25.6) | 15 (17.4) | 7 (8.1) | 4 (4.7) | 2 (2.3) |
| Q3b. | invite stakeholders (i.e., residents, faculty, physicians, community members) to participate in program evaluation(s). | 90 | 0 (0) | 4 (4.4) | 22 (24.4) | 32 (35.6) | 32 (35.6) | 0 (0) |
| Q3c. | clarify the purpose(s) of program evaluations. | 89 | 2 (2.2) | 9 (10.1) | 31 (34.8) | 28 (31.5) | 16 (18.0) | 3 (3.4) |
| Q3d. | develop program evaluation questions. | 89 | 4 (4.5) | 22 (24.7) | 33 (37.1) | 19 (21.3) | 10 (11.2) | 1 (1.1) |
| Q3e. | clarify program logic (i.e., how a program intends to work). | 89 | 3 (3.4) | 14 (15.7) | 34 (38.2) | 23 (25.8) | 13 (14.6) | 2 (2.2) |
| Q3f. | design a program evaluation proposal(s)/plan(s). | 89 | 10 (11.2) | 22 (24.7) | 33 (37.1) | 15 (16.9) | 8 (9.0) | 1 (1.1) |
| Q3g. | design data collection tools for program evaluation (e.g., surveys, interview guides). | 89 | 13 (14.6) | 16 (18.0) | 35 (39.3) | 15 (16.9) | 7 (7.9) | 3 (3.4) |
| Q3h. | collect quantitative program evaluation data. | 89 | 8 (9.0) | 21 (23.6) | 28 (31.5) | 20 (22.5) | 11 (12.4) | 1 (1.1) |
| Q3i. | analyze quantitative program evaluation data. | 89 | 9 (10.1) | 24 (27.0) | 28 (31.5) | 16 (18.0) | 10 (11.2) | 2 (2.2) |
| Q3j. | interpret quantitative program evaluation data. | 88 | 7 (8.0) | 25 (28.4) | 24 (27.3) | 19 (21.6) | 11 (12.5) | 2 (2.3) |
| Q3k. | collect qualitative program evaluation data. | 87 | 4 (4.6) | 9 (10.3) | 36 (41.4) | 16 (18.4) | 20 (23.0) | 2 (2.3) |
| Q3l. | analyze qualitative program evaluation data. | 88 | 4 (4.5) | 14 (15.9) | 35 (39.8) | 18 (20.5) | 15 (17.0) | 2 (2.3) |
| Q3m. | interpret qualitative program evaluation data. | 87 | 4 (4.6) | 13 (14.9) | 34 (39.1) | 18 (20.7) | 15 (17.2) | 3 (3.4) |
| Q3n. | use program evaluation findings. | 87 | 1 (1.1) | 4 (4.6) | 27 (31.0) | 31 (35.6) | 24 (27.6) | 0 (0) |

*Note.* CBD = Competence by Design. Of the 127 respondents who indicated that they engage in program evaluation of their CBD program, between 86 and 90 elected to answer this survey question. Also, percentages do not total to 100 due to rounding.

### *Research Question 2: Why Are CBD Programs Engaging or Not Engaging in Program Evaluation?*

**Survey Question 4a–k: Program Directors' Perceptions About the Purpose of Their CBD Program Evaluation.** Table 5 highlights English and French program directors' perceptions about the purpose of their CBD program evaluation (question 4a–k). Of the 127 respondents, between 85 and 86 elected to answer various parts of this question. Most of the respondents identified that they "frequently" (*n* = 56/86, 65.1%) or "always" (*n* = 13/86, 15.1%) evaluate the needs of their program. The majority (*n* = 54/86, 62.8%) of respondents also indicated that they "frequently" evaluate whether the program addresses its learners' needs. More than three quarters of the respondents indicated that they "sometimes" (*n* = 36/86, 41.9%) or "frequently" (*n* = 32/86, 37.2%) evaluate whether their program has been implemented as intended. Finally, most of the respondents indicated that they "never" (*n* = 38/86, 44.2%) or "rarely" (*n* = 20/86, 23.3%) evaluate the program's cost and cost-effectiveness.

All 127 respondents who indicated that they had engaged in program evaluation of their CBD program were asked an open-ended question about the purpose for which they use program evaluation. Only seven respondents answered the open-ended question, but they each provided multiple explanations about using program evaluation to evaluate their CBD programs for a variety of reasons. The most frequent emerging themes about the reasons for engaging or not engaging in program evaluation concerned informal internal reviews, external accreditation, faculty satisfaction, community needs, and barriers. Of the seven respondents, four (57.1%) shared that they engage in program evaluation for informal internal reviews (i.e., completed by the PGME), two (28.6%) elaborated on how they engage in program evaluation for external accreditation reasons (i.e., completed by the RCPSC), one (14.3%) explained that they engage in program evaluation to meet faculty satisfaction, two (28.6%) noted that they engage in program evaluation to meet the community needs across various training sites, and two (28.6%) emphasized certain barriers (i.e., lack of time and funding) as problematic reasons for not engaging in program evaluation.

**Survey Question 5a–i: Program Directors' Perceptions About Accountability and Learning as Reasons for Program Evaluation.** Table 6 specifies English and French program directors' perceptions about accountability and learning as reasons for program evaluation (question 5a–i). Of the 127 respondents, between 81 and 82 elected to answer various parts of

this question. More than three quarters of respondents indicated that they either "frequently" (*n* = 45/82, 54.9%) or "always" (*n* = 20/82, 24.4%) engage in program evaluation to make decisions about their CBD program. Half of the respondents indicated that they "frequently" (*n* = 41/82, 50.0%) engage in program evaluation to stimulate changes in educational practice. A majority of the respondents indicated that they "frequently" (*n* = 38/82, 46.3%) engage in program evaluation to improve the program's overall practices.

All 127 respondents who indicated that they had engaged in program evaluation of their CBD program were asked an open-ended question about their perceptions regarding accountability and learning as reasons for program evaluation. Only six respondents answered the open-ended question, but they each provided multiple explanations about using program evaluation to evaluate their CBD programs for a variety of accountability and learning reasons. The predominant emerging themes were quality improvement and common barriers. Of the six respondents, three (50.0%) explained that quality improvement is applied to address to problems immediately rather than conducting an exhaustive evaluation and changing CBD implementation, two (33.3%) elaborated on the common barriers of doing program evaluation for learning (e.g., requires too much money, expertise, and time), and one (16.7%) was unsure how to respond to the question.

**Survey Question 6a–h: Program Directors' Perceptions About Different Types of Use as Reasons for Program Evaluation.** Table 7 presents English and French program directors' perceptions about conceptual, instrumental, symbolic, and process use as reasons for program evaluation (question 6a–h). Of the 127 respondents, between 79 and 80 elected to answer various parts of this question. Over three quarters of respondents indicated that they "agree" (*n* = 43/79, 54.4%) or "strongly agree" (*n* = 19/79, 24.1%) their program uses program evaluation to justify decisions previously made about its functioning. Finally, almost all respondents indicated that they "agree" (*n* = 49/80, 61.3%) or "strongly agree" (*n* = 19/80, 23.8%) the behaviour and thinking of their program changes as a result of engaging in program evaluation.

**Table 5**

*Program Directors' Perceptions About the Purpose of Their CBD Program Evaluation*

| | Evaluation purpose | *n* (%) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Survey respondents . . . | *N* | Never | Rarely | Sometimes | Frequently | Always | Don't know |
| Q4a. | evaluate the needs of your program. | 86 | 0 (0) | 3 (3.5) | 14 (16.3) | 56 (65.1) | 13 (15.1) | 0 (0) |
| Q4b. | evaluate if the program addresses its learners' needs. | 86 | 0 (0) | 2 (2.3) | 13 (15.1) | 54 (62.8) | 17 (19.8) | 0 (0) |
| Q4c. | evaluate if the program addresses its faculty's needs. | 86 | 5 (5.8) | 19 (22.1) | 28 (32.6) | 29 (33.7) | 4 (4.7) | 1 (1.2) |
| Q4d. | evaluate how well your program is designed. | 85 | 4 (4.7) | 6 (7.1) | 27 (31.8) | 37 (43.5) | 11 (12.9) | 0 (0) |
| Q4e. | evaluate how well your program is operating. | 86 | 1 (1.2) | 4 (4.7) | 21 (24.4) | 47 (54.7) | 13 (15.1) | 0 (0) |
| Q4f. | evaluate if your program is implemented as intended. | 86 | 1 (1.2) | 8 (9.3) | 36 (41.9) | 32 (37.2) | 8 (9.3) | 1 (1.2) |
| Q4g. | evaluate if learners are satisfied with the program. | 85 | 0 (0) | 1 (1.2) | 11 (12.9) | 45 (52.9) | 28 (32.9) | 0 (0) |
| Q4h. | evaluate if faculty are satisfied with the program. | 86 | 5 (5.8) | 18 (20.9) | 34 (39.5) | 26 (30.2) | 2 (2.3) | 1 (1.2) |
| Q4i. | evaluate your program's outcomes and impact. | 86 | 9 (10.5) | 9 (10.5) | 32 (37.2) | 30 (34.9) | 5 (5.8) | 1 (1.2) |
| Q4j. | evaluate the program's cost and cost-effectiveness. | 86 | 38 (44.2) | 20 (23.3) | 13 (15.1) | 11 (12.8) | 2 (2.3) | 2 (2.3) |
| Q4k. | investigate if program resources are used efficiently. | 86 | 20 (23.3) | 21 (24.4) | 25 (29.1) | 15 (17.4) | 4 (4.7) | 1 (1.2) |

*Note.* CBD = Competence by Design. Of the 127 respondents who indicated that they engage in program evaluation of their CBD program, between 85 and 86 elected to answer this survey question. Also, percentages do not total to 100 due to rounding.

**Table 6**

*Program Directors' Perceptions About Accountability and Learning as Reasons for Program Evaluation*

| | Accountability and learning as reasons | *n* (%) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Survey respondents engage in program evaluation to . . . | *N* | Never | Rarely | Sometimes | Frequently | Always | Don't know |
| Q5a. | satisfy accountability demands. | 81 | 11 (13.6) | 16 (19.8) | 23 (28.4) | 18 (22.2) | 6 (7.4) | 7 (8.6) |
| Q5b. | produce evidence of program effectiveness. | 81 | 12 (14.8) | 17 (21.0) | 29 (35.8) | 17 (21.0) | 2 (2.5) | 4 (4.9) |
| Q5c. | make decisions about the program. | 82 | 0 (0) | 2 (2.4) | 15 (18.3) | 45 (54.9) | 20 (24.4) | 0 (0) |
| Q5d. | develop evaluation and inquiry skills. | 82 | 13 (15.9) | 19 (23.2) | 29 (35.4) | 14 (17.1) | 4 (4.9) | 3 (3.7) |
| Q5e. | understand programming issues more fully. | 82 | 5 (6.1) | 15 (18.3) | 31 (37.8) | 18 (22.0) | 9 (11.0) | 4 (4.9) |
| Q5f. | stimulate changes in clinical practice. | 82 | 11 (13.4) | 15 (18.3) | 28 (34.1) | 22 (26.8) | 4 (4.9) | 2 (2.4) |
| Q5g. | stimulate changes in educational practice. | 82 | 1 (1.2) | 8 (9.8) | 22 (26.8) | 41 (50.0) | 10 (12.2) | 0 (0) |
| Q5h. | learn about other people's perspectives of programs. | 82 | 10 (12.2) | 10 (12.2) | 18 (22.0) | 35 (42.7) | 7 (8.5) | 2 (2.4) |
| Q5i. | improve its overall practices. | 82 | 2 (2.4) | 9 (11.0) | 23 (28.0) | 38 (46.3) | 10 (12.2) | 0 (0) |

*Note.* Of the 127 respondents who indicated that they engage in program evaluation of their CBD program, between 81 and 82 elected to answer this survey question. Also, percentages do not total to 100 due to rounding.

**Table 7**

*Program Directors' Perceptions About Different Types of Use as Reasons for Program Evaluation*

| | Uses for program evaluation | n (%) | | | | | |
|---|---|---|---|---|---|---|---|
| | Survey respondents' program . . . | *N* | Strongly disagree | Disagree | Agree | Strongly agree | Don't know |
| Q6a. | uses program evaluation to better understand its functioning and practices. | 80 | 0 (0) | 6 (7.5) | 51 (63.8) | 22 (27.5) | 1 (1.3) |
| Q6b. | uses program evaluation to learn from its experiences. | 80 | 0 (0) | 3 (3.8) | 48 (60.0) | 28 (35.0) | 1 (1.3) |
| Q6c. | bases decisions about functioning on program evaluation results. | 80 | 0 (0) | 9 (11.3) | 47 (58.8) | 22 (27.5) | 2 (2.5) |
| Q6d. | uses program evaluation results to influence changes in program policies and procedures. | 80 | 1 (1.3) | 7 (8.8) | 46 (57.5) | 24 (30.0) | 2 (2.5) |
| Q6e. | uses program evaluation to comply with reporting demands. | 80 | 3 (3.8) | 14 (17.5) | 34 (42.5) | 25 (31.3) | 4 (5.0) |
| Q6f. | uses program evaluation to justify decisions previously made about its functioning. | 79 | 2 (2.5) | 12 (15.2) | 43 (54.4) | 19 (24.1) | 3 (3.8) |
| Q6g. | uses program evaluation to develop its research and inquiry skills. | 80 | 11 (13.8) | 23 (28.8) | 33 (41.3) | 9 (11.3) | 4 (5) |
| Q6h. | behaviour and thinking changes as a result of engaging in program evaluation. | 80 | 0 (0) | 9 (11.3) | 49 (61.3) | 19 (23.8) | 3 (3.8) |

*Note.* Of the 127 respondents who indicated that they engage in program evaluation of their CBD program, between 79 and 80 elected to answer this survey question. Also, percentages do not total to 100 due to rounding.

All 127 respondents who indicated that they had engaged in program evaluation of their CBD program were asked an open-ended question about their uses for program evaluation. Only four respondents answered this open-ended question, but they each provided multiple explanations about evaluating their CBD programs for a variety of uses. The predominant emerging themes were contributions to scholarly CBD conversations and clarification on a program's lack of agency. Out of the four respondents, one (25.0%) explained that they or others within their program use program evaluation to contribute to scholarly conversations on several topics related to CBD, and two (50.0%) explained that their program does not have agency, so their program's behaviour or thinking cannot change. Although feedback from the internal and external program evaluation reviews allow for further information sharing, this sharing may or may not stimulate program change. Stimulation for change comes from the PGME and the RCPSC levels, which hold the *power* to change a program's behaviour or thinking. Finally, one (25.0%) explained that they were unsure about the terminology of "program evaluation," but advocate that they constantly evaluate their program through quarterly committee discussions.

**Survey Question 7: Program Directors' Response to Whether Their Program has an Employee Whose Primary Responsibility is Program Evaluation.** Table 8 specifies English and French program directors identifying whether their program has an employee whose primary responsibility is program evaluation (question 7). Of the 80 respondents who answered this question, 69 (86.3%) indicated that "no" their program does not have an employee whose primary responsibility is program evaluation.

**Survey Question 8: Program Directors' Response to Whether Their Program Receives Funding for Program Evaluation.** Table 9 specifies English and French program directors identifying whether their program receives funding for program evaluation (question 8). Of the 81 respondents who answered this question, 64 (79.0%) indicated that "no" their program does not receive funding for program evaluation.

**Table 8**

*Program has an Employee Whose Primary Responsibility is Program Evaluation*

| Employee primary responsibility | | *n* (%) | | | |
|---|---|---|---|---|---|
| | | *N* | Yes | No | I don't know |
| Q7. | Survey respondents whose program has an employee whose primary responsibility is program evaluation. | 80 | 7 (8.8) | 69 (86.3) | 4 (5) |

*Note.* Of the 127 respondents who indicated that they engage in program evaluation of their CBD program, 80 elected to answer this survey question. Also, percentages do not total to 100 due to rounding.

**Table 9**

*Program Receives Funding for Program Evaluation*

| Funding | | *n* (%) | | | | |
|---|---|---|---|---|---|---|
| | | *N* | Yes | No | I don't know | I prefer not to say |
| Q8. | Survey respondents whose program receives funding for program evaluation. | 81 | 5 (6.2) | 64 (79.0) | 11 (13.6) | 1 (1.2) |

*Note.* Of the 127 respondents who indicated that they engage in program evaluation of their CBD program, 81 elected to answer this survey question.

**Survey Question 9a–d: Program Directors' Perceptions About Reasons That Their Program does not Engage in Program Evaluation.** Table 10 specifies English and French program directors identifying reasons their program does not engage in program evaluation (question 9a–d). After question 1 (i.e., consenting to survey participation) and question 2 (i.e., selecting "does not engage in program evaluation"), respondents skipped questions 3-8 and were branched to question 9. Of the 16 respondents who indicated that their CBD program does not engage in program evaluation, 13 elected to answer questions about the reasons for not engaging in program evaluation. Respondents were permitted to select all of the reasons that applied. Out of the 13 respondents, 7 (53.8%) indicated that their program does not engage in program evaluation because the program has no personnel to do program evaluation, 7 (53.8%) indicated that their program does not engage in program evaluation because the program has no funding to do program evaluation, and a majority of respondents ($n$ = 9, 69.2%) indicated that their program

does not engage in program evaluation because the program does not know how to do program evaluation.

**Table 10**

*Reasons Program does not Engage in Program Evaluation*

| | Not engaging in program evaluation | | *n* (%) |
|---|---|---|---|
| | Survey respondents whose program does not engage in program evaluation because . . . | *N* | Selected |
| Q9a. | the program has no personnel to do program evaluation. | 13 | 7 (53.8) |
| Q9b. | the program has no funding to do program evaluation. | 13 | 7 (53.8) |
| Q9c. | the program has no time to do program evaluation. | 13 | 4 (30.8) |
| Q9d. | the program does not know how to do program evaluation. | 13 | 9 (69.2) |

*Note.* Of the 16 respondents who indicated they do not engage in program evaluation of their CBD program, 13 elected to answer this survey question. Respondents were permitted to select all that apply; thus the percentage column's sum is greater than 100 because respondents could provide more than one response.

All 16 respondents that noted that their CBD program does not engage in program evaluation were asked an open-ended question to identify any other reasons why their program does not engage in program evaluation. Only six respondents answered this open-ended question, but they each provided multiple explanations about various reasons for not using program evaluation to evaluate their CBD programs. The predominant emerging themes were that program evaluation is not used to evaluate their CBD programs because they are unsure about the meaning of "program evaluation" and they instead engage in informal evaluation as a form of quality improvement. Of the six respondents, two (33.3%) elaborated that they do not engage in program evaluation because they are unsure of what is meant by program evaluation, and three (50.0%) shared that they do not engage in program evaluation because they complete informal assessments to address issues as they occur, but they do not apply a systematic process involving an external evaluator and a standardized approach. Finally, one (16.7%) clarified that their program's transition to CBD was delayed by a year and, thus, have yet to evaluate it.

**Survey Question 10a–b: Program Directors Interested in Receiving a Summary of the Survey Results.** Table 11 presents English and French program directors' interest in receiving a summary of the survey results (question 10a–b). Of the 93 respondents who answered this question, 60 (64.5%) indicated "yes" they are interested in receiving a summary of the survey results, and of those 60 respondents, 57 (95.0%) provided their email addresses.

**Table 11**

*Program Directors Interested in Receiving a Summary of the Survey Results*

| Interest in receiving a summary of the survey results | | n (%) |
|---|---|---|
| | *N* | Selected |
| Q10a. "Yes" from respondent | 93 | 60 (64.5) |
| "No" from respondent | 93 | 33 (35.5) |
| Q10b. After clicking "yes," respondent provided their email address | 60 | 57 (95.0) |

*Note.* Of the 149 respondents who consented to completing this survey, 93 elected to answer this question.

**Summary for Phase 1**

Overall, 149 program directors from Canadian PGME specialty and subspecialty programs who have transitioned to CBD consented to participate in Phase 1 of this study. Of the 149, 143 respondents specified whether their program engages or does not engage in program evaluation. Over three quarters (*n* = 127/143, 88.8%) of respondents indicated that their program does engage in program evaluation, and 16 (11.2%) indicated that their program does not engage in program evaluation. Of the respondents whose CBD programs engage in program evaluation, more than half of the respondents indicated that they have either "never" (*n* = 36/86, 41.9%) or "rarely" (*n* = 22/86, 25.6%) worked with a program evaluator. Most of the respondents indicated that they invite stakeholders to participate in program evaluations either "frequently" (*n* = 32/90, 35.6%) or "always" (*n* = 32/90, 35.6%). More than half of the respondents answered that they "rarely" (*n* = 25/88, 28.4%) or "sometimes" (*n* = 24/88, 27.3%) interpret quantitative program evaluation data. Furthermore, when reflecting about the purpose of their CBD program evaluation more than three quarters of the respondents indicated that they "sometimes" (*n* = 36/86, 41.9%) or "frequently" (*n* = 32/86, 37.2%) evaluate whether their program has been implemented as intended. However, most of the respondents indicated that they "never" (*n* = 38/86, 44.2%) or "rarely" (*n* = 20/86, 23.3%) evaluate the program's cost and cost-effectiveness. When considering accountability and learning as reasons for program evaluation, more than three quarters of respondents indicated that they either "frequently" (*n* = 45/82, 54.9%) or "always" (*n* = 20/82, 24.4%) engage in program evaluation to make decisions about their CBD program. Most respondents (*n* = 69/80, 86.3%) indicated that "no" their program does not have

an employee whose primary responsibility is program evaluation. Most respondents (*n* = 64/81, 79.0%) indicated that "no" their program does not receive funding for program evaluation.

The remaining respondents (*n* = 16/143, 11.2%) indicated that their CBD programs do not engage in program evaluation. Most of these respondents (*n* = 13/16, 81.3%) indicated the reasons for not engaging in program evaluation and were permitted to select all the reasons that applied. Over half of respondents (*n* = 7/13, 53.8%) indicated that their program does not engage in program evaluation because the program has no personnel to do program evaluation. Similarly, over half of the respondents (*n* = 7/13, 53.8%) indicated that their program does not engage in program evaluation because the program has no funding to do program evaluation, and a majority of respondents (*n* = 9/13, 69.2%) indicated that their program does not engage in program evaluation because the program does not know how to do program evaluation.

Respondents also indicated that engagement in program evaluation is limited due to cost, time, and expertise constraints. The collection, analysis, and evaluation of program data are low priorities because program directors lack time. Moreover, when program evaluation findings are acquired, there is a lack of power to make changes because the PGME office and the RCPSC make decisions. Thus, some program director stakeholders do not feel empowered to engage in program evaluation. Furthermore, some of the suggested program revisions are not possible because of budget constraints. Some program directors indicated that they are unclear about what program evaluation means. Finally, certain program directors indicated that they had started to assess their programs in an attempt to encourage ongoing quality improvement, but feel that a standardized approach is required.

These Phase 1 findings informed the Phase 2 interview guide to support answering the Phase 2 questions. Specifically, in the Phase 1 open-ended responses, respondents suggested that they and other individuals engage in program evaluation during internal and external reviews. An internal review of residency programs is conducted by the PGME office mid-accreditation cycle. An external review is also performed by the RCPSC accreditation visit. Respondents often equated program evaluation with continuous or ongoing quality improvement as labelled by the RCPSC. My instrument guide prompted further into what this looks like in their PGME environment. Additionally, some respondents indicated that they may evaluate their program, but not in the systematic way, involving an external evaluator. Instead, program directors complete informal assessments of the program and address issues as they arise. My Phase 2 interview

guide explored how program directors engage in this informal program evaluation and their suggestions for building capacity to develop systematic evaluation. Similarly, my Phase 3 interview guide prompted program evaluators to learn more about how they can assist with establishing formalized program evaluation in PGME departments.

# Chapter 5: Phase 2

In this chapter, I describe the methods and results of Phase 2, during which I investigated program directors' perspectives about how they or others in their CBD program were engaging in program evaluation and how they or others in their CBD program could build their capacities to do program evaluation and use evaluation findings. For this component of the study, I interviewed the program directors who, on the Phase 1 survey, had indicated interest in participating in a follow-up interview. The purpose of this phase was to explore the following research questions:

- Question 3: How are CBD programs engaging in program evaluation?
- Question 4: How can CBD programs build their capacities to do program evaluation and use evaluation findings?

## Sample

I applied convenient sampling (i.e., non-probability sampling where the researcher collects data from a conveniently available group of respondents) to identify and recruit program directors whose CBD programs either engage or do not engage in program evaluation, who had consented and completed the survey in Phase 1. Of the 149 respondents, 36 (24.2%) shared their email addresses to be contacted about participating in a follow-up interview. Of the 36 program directors who agreed to be contacted to participate in a follow-up interview, 15 consented to be interviewed for Phase 2.

## Instrument Development

I used the findings from Phase 1 to design the semi-structured interview guides for Phase 2 (see Appendix H). Tables 12 and 13 present the research questions that guided Phase 2 and the corresponding interview guide questions. I created separate guides for respondents who indicated that they engage (Table 12) or do not engage (Table 13) in program evaluation to evaluate their CBD program. I began both guides with an introductory script and included open-ended questions with additional probes to allow for expansion and idea prompting. The guide for those who engage in program evaluation included ten questions, while the guide for those who do not engage in program evaluation included nine questions. I piloted the guides with my supervisor and another program evaluation researcher to ensure they were clearly worded and included relevant questions or probes (Seidman, 2019).

**Table 12**

*Specifications for Phase 2 Interview Guide: Engages in Program Evaluation*

| Research questions | Corresponding interview questions |
|---|---|
| 3. How are CBD programs engaging in program evaluation? | Question 1<br>Question 2 |
| 4a. How can CBD programs build their capacities to do program evaluation? | Question 3<br>Question 4<br>Question 5<br>Question 6<br>Question 7<br>Question 8 |
| 4b. How can CBD programs build their capacities to use evaluation findings? | Question 9<br>Question 10 |

*Note*. CBD = Competence by Design.

**Table 13**

*Specifications for Phase 2 Interview Guide: Does not Engage in Program Evaluation*

| Research questions | Corresponding interview questions |
|---|---|
| 3. How are CBD programs engaging in program evaluation? | Question 1<br>Question 2<br>Question 3 |
| 4a. How can CBD programs build their capacities to do program evaluation? | Question 4<br>Question 5<br>Question 6<br>Question 7 |
| 4b. How can CBD programs build their capacities to use evaluation findings? | Question 8<br>Question 9 |

*Note*. CBD = Competence by Design.

## Data Collection Procedures

I emailed the eligible 36 program directors a letter of information and a consent form (see Appendix I and Appendix J) inviting them to participate in an interview for Phase 2. The information letter asked the program directors to reply to me if they were interested in participating. Each interview was scheduled at a time that was convenient for the interviewee, took place using Zoom (https://zoom.us), and lasted approximately one hour. Prior to the

interview, all participants signed and returned an informed consent form to me. I recorded and transcribed each interview verbatim. Each transcription was sent to the corresponding interviewee for approval and feedback. I applied any changes identified by the interviewees. I gave each participant a $10 Amazon gift card to thank them for participating in the interview.

## Data Analysis

I followed Miles et al.'s (2019) systematic step-by-step approach to acquire meaning from data and confirm findings. First, my supervisor and I analyzed the data by tracking emerging themes from the interviews, beginning with deductive codes based on my literature review. We used NVivo software to organize the data. Next, we refined the coding system by including inductive codes to generate a comprehensive and reflective code list (see Appendix K for the final Phase 2 NVivo codes). We used the themes, codes, and coded quotations to create a data analysis matrix, which we used to highlight and understand the relationships between themes and codes, and to make note of irregular or contradictory findings (see Appendix L for a sample matrix). Each interviewee was identified using a pseudonym made up of the letter "P" for program director and a number based on the interview order (e.g., P1, P2, etc.). I then invited all the interviewees to review the preliminary study findings and offer their perspectives (Creswell, 2014).

Finally, in my qualitative analysis, I applied Patton's (2015a) suggestions to avoid quantifying my qualitative data. Specifically, I intentionally decided to not quantify or count the number of participants who mentioned ideas that I associated with representing each theme. In this way, I followed Patton's (2015a) arguments to remain qualitative because "it's not how many said something that matters" (p. 557). Therefore, I have omitted the number of respondents who shared each theme, so I can emphasize the participants' message content and not take away from their perspectives. Instead of reporting all quotations collected from participants, I have provided exemplar quotations to establish a "verbal display that represents and presents data vividly about the study's phenomenon of interest" (Miles et al., 2019, p. 324). I have selected these example quotes because they accurately highlight the interviewee perspectives about engaging in evaluation of a CBD program.

### *Trustworthiness*

I used the following three criteria to establish the trustworthiness of my qualitative research and to ensure that the findings validated the experiences of the individuals under

investigation: (a) credibility (e.g., the appropriateness and accuracy of the data sources as well as the interpretation of the data), (b) confirmability (e.g., the congruence between two or more independent people about the accuracy, relevance, or meaning of the study; Lincoln & Guba, 1985), and (c) transferability (e.g., whether the findings from the data can be transferred to other settings or groups; Lincoln & Guba, 1985). First, I used credibility to evaluate trustworthiness by member-checking (i.e., participant verification of the data's accuracy in representing their perceptions and experiences; Lincoln & Guba, 1985). Following the interview and coding of qualitative data using NVivo, I verified my summary of the main and reoccurring ideas with the interviewees to ensure accuracy of results. Second, I used confirmability to ensure trustworthiness in my qualitative data by comparing the interviewee main ideas in NVivo to find similar or diverging data to identify meaning from the study. These congruencies and discrepancies inform my discussion in Chapter 7. Finally, I used transferability to evaluate trustworthiness by comparing whether similar findings have been published in journals about research on program evaluation. Furthermore, I used notetaking to document my assumptions, questions, and decisions during the data interpretation. This descriptive notetaking strengthens the trustworthiness of the findings by making this study replicable in other settings.

## Findings

### *Characteristics of Participants*

Of the 15 interviewees, 14 (93.3%) indicated that they engage in program evaluation of their CBD program, while the remaining one (6.7%) indicated that they do not. The participants included program directors and associate program directors from a variety of specialty and subspecialty Canadian PGME programs.

**Table 14**

*Phase 2: Theme Findings from Program Director Interviews*

| Research Question | Theme Findings |
|---|---|
| ***3: How are CBD programs engaging in program evaluation?*** | (a) Struggling to engage in program evaluation because of limited available resources and buy-in<br>(b) Using ad hoc evaluation methods<br>(c) Using a team-based evaluation format |
| ***4a: How can CBD programs build their capacities to do program evaluation?*** | (a) Developing expertise in their program evaluation abilities<br>(b) Acquiring program evaluation resources for guidance<br>(c) Advocating for clear program evaluation expectations from the RCPSC, the PGME office, and within their specialties |
| ***4b: How can CBD programs build their capacities to use evaluation findings?*** | (a) Asking meaningful questions that are appropriate to the stage of the program<br>(b) Establishing buy-in for program evaluation to increase the likeliness of using the findings<br>(c) Sharing findings with faculty, residents, and other programs in ways that encourage use<br>(d) Understanding evaluation as helpful for the accreditation process so program directors are more likely to use the findings |

### *Research Question 3: How Are CBD Programs Engaging in Program Evaluation?*

In the sections that follow, I describe how the participants' CBD programs are engaging in program evaluation. Specifically, the programs are: (a) struggling to engage in program evaluation because of limited available resources and buy-in; (b) using ad hoc evaluation methods; and (c) using a team-based evaluation format.

**Struggling to Engage in Program Evaluation Because of Limited Available Resources and Buy-In.** Participants explained that their CBD programs are engaging in program evaluation by balancing their limited resources with limited stakeholder buy-in. Program directors are having difficulty doing evaluation within their specialties/subspecialties due to limited financial support, human resources, and time. Participants stated that "the simple answer [is] it's time and money. Unfortunately" (P1), because "CBD came in, but no extra money came in. And, so, I think programs in general have found that CBD costs a lot more than

what the Royal College was suggesting it would" (P3). These costs mean that faculty are required to implement CBD without additional resources to evaluate the actual implementation. As such, "people are needing to figure out ways to do more with the same amount of resources, but without help" (P3). According to these participants, the minimal program evaluation on CBD implementation stems from a lack of additional funding to allocate human resource support and the already stretched time of busy faculty members. Faculty members who are involved with committees are volunteering their time to support program evaluation, but their demanding schedules means their time is limited. Participants explained that "everybody else in our training committee, it's completely volunteer basically and . . . everybody's just got too many things to do" (P12).

Without additional resources, faculty are hesitant to add more to their workload demands. Participants clarified that "it's just not realistic for people who are unpaid volunteers, working full time as physicians to do this kind of work" (P12), which makes stakeholder buy-in difficult because it involves "asking our staff to take more time out to do the evaluations. It's been really challenging" (P11). Participants also emphasized that "everyone is very busy and reluctant to add more to their plate" (P12) and faculty believe that informal program evaluation is already being completed because "we talk about what's going on, what's working well, what's not working well. Why do we have to do any more work for it?" (P12). Moreover, participants indicated that the varying factors involved in residency medical education would make it "very hard to prove with program evaluation that CBD as a whole has improved residency education" (P13) and that they "wonder if there's still this much in flux. Whether it's useful to start a program evaluation right away?" (P7). They highlighted that it is not only faculty, but also that "residents [are] just too burned out from surveys because of other issues" (P6). Finally, participants identified the challenge with getting buy-in from faculty and residents to engage in program evaluation and how it is reflected in the quality of the evaluation. Specifically, "the resources, time, remuneration, people are just inadequate . . . It's patchwork" (P14).

In summary, participants discussed how some CBD programs are struggling to do program evaluation because of limited available resources and challenges associated with securing stakeholder buy-in.

**Using Ad Hoc Evaluation Methods.** Participants indicated that their CBD programs are engaging in program evaluation by using ad hoc methods without appropriate forethought and

follow up. They expressed feelings of uncertainty when creating data collection tools or a base of comparison for program evaluations. For example, a program director stated, “I don’t feel like there’s a validated tool or benchmark of what we should be looking at” (P15). This lack of appropriate tools and information leads to programs improvising methods and processes, including the use of experiential information, surveys, informal feedback, and exit interviews. Participants indicated, “we’re not super confident. We’re improvising” (P15). Participants explained that “at this point, we’re linking [program evaluation] to our surveys and our informal feedback . . . but, I think there can definitely be [a] more robust and fulsome mechanism for providing that” (P10). In addition, participants described how “it’s a bit of improvising the program evaluation” using “homemade surveys” (P15) or “it’s just myself and the program assistant who are designing the surveys that we solicit this information from the residents directly” (P10). Consequently, a lack of confidence in the evaluation process develops because “there’s not a lot of insight that can be gained from probably some of the surveys that I can design” (P10). Without formalizing the program evaluation, participants noted that “[programs] can’t really target or pinpoint what specific interventions do . . . [and are] going on the assumption that if we’re improving, the changes we are making over time are working” (P13).

In summary, participants discussed how some CBD programs are engaging in program evaluation using ad hoc evaluation methods that lack appropriate forethought and follow up.

**Using a Team-Based Evaluation Format.** Participants specified that their CBD programs are engaging in program evaluation using a team-based format with varying degrees of involvement from PGME programs. For example, participants indicated that there is evaluation participation from “postgraduate, university-level leadership . . . where we actually help review our new CBME curriculum and . . . get the input from the different sites where residents rotate” (P15). Within certain larger program specialties, the program directors also reported that they have an associate program director or multiple faculty members who help with program evaluation. They explained that during their experience as “associate program director, there was another associate program director. There were two of us in that role who could do some evaluation” (P6). In addition, participants noted that there are committees that discuss program evaluation activities, including a “CBME designated committee [who] meet on two-month basis [with] representatives from all of the program leadership, and then the residents” (P15).

Furthermore, participants indicated that their CBD programs are also working with administrative members to support data analysis and to share evaluation findings within and across universities, as well as with program evaluation experts in some cases, to learn more about the evaluation process. For example, some programs "asked that all program feedback go to our PA, our program admin, who compiled comments—whether they were good or bad" (P5). Participants indicated that they shared findings with "subspecialties [who] are just coming online now. So, [we shared] a lot of data learned from our process and what we've kind of been working through and struggling with" (P13). In addition to sharing within universities, participants indicated that there is cooperative learning and evaluation across universities. This learning occurred at a "once or twice a year meeting of all the [Canadian specialty] program members . . . and we set up a platform where people can upload stuff that might be of use to other programs" (P12). However, despite collaboration among certain programs, participants indicated that they "don't know what other programs are doing" (P10) and that they need support from "organizations to help with evaluation implementation projects" (P11). At the university level, participants noted that "a program evaluation expert from the PGME office . . . helped a lot [with] devising basically evaluation questions, and I'm not great with the terminology, but like a logic model" (P12). Specifically, participants described how they "started out with a rough idea, and what I thought we should evaluate and showed that to her and then she looked at it and basically helped to formalize it a little bit more" (P12).

In summary, participants discussed how some CBD programs are engaging in program evaluation using a team-based format with varying degrees of input from their PGME programs.

***Research Question 4a: How Can CBD Programs Build Their Capacities to Do Program Evaluation?***

In the sections that follow, I describe how the participants' CBD programs that are or are not engaging in program evaluation can build their capacities to do program evaluation. Specifically, the programs can: (a) develop expertise in their program evaluation abilities; (b) acquire program evaluation resources for guidance; and (c) advocate for clear program evaluation expectations from the RCPSC, the PGME office, and within their specialties.

**Develop Expertise in Their Program Evaluation Abilities.** Participants suggested that CBD programs can build their capacities to do program evaluation by increasing their expertise in program evaluation skills. Program directors expressed statements such as, "I'm so unfamiliar

with the process to evaluate the program . . . but the idea sounds very interesting. I think it sounds very useful potentially . . . It sounds like a good way to potentially improve your programming" (P7). They also acknowledged how program evaluation professional development sessions offered an opportunity to develop their expertise in program evaluation. Participants indicated that "we also had some training, a retreat, about program evaluation around the same time. And, so, then we decided, well maybe this is actually a useful thing to do" (P12). They explained that "there are program director retreats where all program directors at the university share ideas and what is working and what is not. There are options/workshops about program evaluation" (P2). Participants also noted that faculty development increased their understanding of program evaluation potential and encouraged explicit questioning, measurement of interventions, and reflection on the intervention's improvement. They expressed an interest in "more training . . . in program evaluation" and training for faculty that focused on establishing "comfort around coaching around the growth mindset . . . in medical education and program evaluation . . . for our learners and our supervisors" (P6).

In summary, participants discussed how developing expertise in their evaluation abilities through workshops, retreats, coaching, and other types of training, is necessary for building their CBD programs' capacities to do program evaluation.

**Acquire Program Evaluation Resources for Guidance.** Participants suggested that CBD programs can build their capacities to do program evaluation through access to program evaluation resources. Some participants indicated that their programs "haven't done [program evaluation] really at this point . . . because [we] don't know where to get started, [we] don't know what type of resources are available to be able to do that" (P10). Participants advocated for the creation of and access to resources. A "how-to" resource should "be streamlined, easy, reproducible, and then you get people [who] actually use it. Lacking any one of those things and it's too much" (P8). These participants indicated that it feels as if they are "rebuilding the wheel. And that's why [we] say it'd be nice to have formalized resources out there because [we] know that there are other programs that do this and do this well" (P10). They also stated that they "would like some resources, some examples that help, or at least structure the programmatic review in a more formalized setting" (P10). Moreover, participants noted that pre-developed tools would yield accurate and informative data to answer the question(s) of interest. As such,

"it's always getting more resources centrally of 'what is it that we're supposed to do?'" (P15). Participants also commented on the idea of including examples.

Participants also suggested the need for funding to add relevant human resources and to allocate more time to build program evaluation capacity. Participants clarified that "more time, more money would be great" (P11) and a need for "humans, people with the right skills and expertise" (P4). As such, "a researcher or research assistant or someone who can come in and kind of has a bit of a clinical sense, but has an eye for okay, 'this is what we need to do, this would be great to look at' . . . would be really helpful" (P11). Participants indicated that "know the data is there. [They] need someone to do a deep dive" (P1) and "it would be useful if post grad programs made available somebody who can help with analysis . . . from the raw data into something usable" (P9). Allocating a research assistant with a clinical background to gather data and complete more administrative tasks would allow program directors to allocate more time to higher level program evaluation planning and analysis. These participants insisted that to build capacity, they need "time, time to do it" (P4) and "what would really be useful is . . . really time. At least for me, that's what it always was" (P12).

In addition, an evaluation expert could provide support to building the capacity to do program evaluation. Participants explained that they "need a spirit guide, resources and examples of . . . matrices and programmatic assessment that would be applicable to our specialty, and it can be used to . . . gather data" (P10). This expert is essential because "there's not a lot of insight that can be gained from probably some of the surveys that [we] can design, but there are other people out there who are much better at designing assessment type of questions that would be helpful" (P10). Although some participants accessed an evaluation expert within their PGME, others were "not aware of an external specific program evaluator—that would be nice" (P11). Participants advocated that they "need . . . someone whose job focus is to be able to have the time to do that deep dive in and look at these questions and [then the program directors are] able to then bring that forward and bring some practicality to it" (P14). They also felt the "need [for] a dedicated individual, that isn't really responsible for caring for patients . . . to evaluate and improve our program implementation" (P14). Participants expressed interest in having "somebody that the PGME provided that was able to evaluate the program and then tell [us] what was working and what wasn't working. That would be fantastic because [we] don't know how to do that" (P3). Although participants acknowledged that funding may not allow for a PGME

expert to evaluate the program and to identify strengths and weaknesses, they suggested that a shared program evaluator among a few programs who has "the expertise in the department would be great—the guidance" (P11), especially for a beginner, "working with a qualified program evaluator can be very, very useful" (P6).

Finally, participants suggested that access to technical resources can help build their CBD programs' capacities to do program evaluation by acquiring information, organizing data, simplifying the data analysis process, and encouraging resource template sharing. Specifically, to access the captured data and learn from it, participants indicated that programs must have adequate "IT infrastructure because that's the key" (P4) and "it would be great if a tool was built to help" (P1). For example, "IT support for Elantra. Trying to get data out of Elantra in a readable format and having a bit more IT support for that . . . as well would be really helpful" (P11). Participants also noted that consistency among Canadian PGME software could support resource evaluation template sharing of data collection tools. However, Canadian PGME programs are "not all on 145 for example. We're kind of limited by what our PGME has decided to go with in terms of software and what's allowed" (P5). Thus, participants suggested that "it would be more useful to have [exemplars and templates] at PGME level than national level because, we find it hard to share across the country" (P5) on various digital platforms.

In summary, participants discussed how acquiring program evaluation resources, such as ready-to-use examples, templates, funding for additional human resources to make more time available, access to program evaluation experts, and access to software and IT support, is essential for building their CBD programs' capacities to do program evaluation.

**Advocate for Clear Program Evaluation Expectations from the RCPSC, the PGME Office, and Within Their Specialties.** Participants suggested that CBD programs can build their capacities to do program evaluation by clarifying expectations. Participants explained that "if the Royal College wants program[s] to have an ability to do program evaluation, [they need] to be explicit about that. Because I've not seen that's necessarily a rigid requirement that they have" (P6). For example, "at a higher level, [the RCPSC] should be saying, 'here are the common questions we think you should all be looking at.' And then also recognize that each program will have unique variations of those questions" (P4). Participants also suggested that policy creation and intentions should be made "either at the Royal College level, or the PGME level or some other level kind of mandating it" (P5), and that the higher policy level should include "more

direction from the Royal College, rather than saying, 'you should engage in program evaluation,' which is basically exactly what they tell us. [However], what does that mean?" (P12).

Participants also expressed the need for clear expectations of what to evaluate from PGME leadership. For example, it could "be at a PGME level or even something like a specialty committee level, where we have a unified front that even could be specific to a specialty which helps with assessment and a standardized or nationalized type of assessment strategy" (P10). Participants noted that they "need to have discussions also with our PGME and our program leadership to determine like, what are [they] wanting to see from our conversion to CBD? What do [they] see as being valuable for us to focus on evaluation?" (P6). Not only is it important for policies that clarify what is being evaluated but it is also necessary to be explicit about who is responsible for leading the evaluation. As such, participants commented that "if the [RCPSC and PGME] policies and procedures for cyclical review incorporate the philosophy of program evaluation, then effectively, the faculty that is taking the lead on doing the cyclical review of that module is learning through doing" (P2).

By advocating for clear expectations from the RCPSC, the PGME, and specialty programs about evaluation, it begins "really formalizing it . . . [thinking about] how are we going to make that decision? What are we going to base that decision off of?" (P13). Program directors noted that "the post grad is working towards a standardized process for programs across the university . . . and I see it coming towards program evaluation too . . . so, if we get more of that [it is] going to be helpful" (P9). Participants suggested that establishing guidelines about program evaluation will help evaluations to "be overall more formal, evidence-based in the sense that there's validity to how we're evaluating the program, and that there's . . . a feedback loop that's there every year or so that feeds into the program development and how it's rolled out" (P15). Thus, there will be "a roadmap or something to follow that we actually know that we're not wasting our time. And that what we're looking at, in terms of the program evaluation, is of value" (P8). To support consistency across Canadian specialty/subspecialty programs, "a standardized roadmap . . . could be very similarly used across the country, so that we're all kind of on the same page" (P8).

In summary, participants discussed how advocating for clear program evaluation expectations from the RCPSC, the PGME office, and within their specialty, can help their CBD programs build their capacities to do program evaluation.

### *Research Question 4b: How Can CBD Programs Build Their Capacities to Use Evaluation Findings?*

In the sections that follow, I describe how the participants' CBD programs that are or are not engaging in program evaluation can build their capacities to use evaluation findings. Specifically, the programs can: (a) ask meaningful questions that are appropriate to the stage of the program; (b) establish buy-in for program evaluation to increase the likeliness of using the findings; (c) share findings with faculty, residents, and other programs in ways that encourage use; and (d) understand evaluation as helpful for the accreditation process so program directors are more likely to use the findings.

**Ask Meaningful Questions That Are Appropriate to the Stage of the Program.** Participants suggested that CBD programs can build their capacities to use evaluation findings by asking meaningful questions that are relevant to the program implementation stage. Participants acknowledged that the different hierarchical levels (i.e., RCPSC, PGME, and program specialty/subspecialty) will have different questions that are meaningful to them: "depending on what level you do your evaluation at, the questions will be different? Central, PGME, if you're looking at the Royal College" (P4). Participants also suggested questions that could encourage use. For example, participants asked "is there any way to measure how CBME impacts resident training? . . . [we] don't really have an answer. So that's why we're trying to evaluate a little bit" (P15). Whereas other participants noted "one area that I think would be interesting is . . . identifying the residents in difficulty" (P4) and "looking to see with CBD, [are] there fewer delays in training? [Are] there fewer probations? Looking at kind of those end points" (P13). Capturing evidence by "looking for some sort of event rate . . . wouldn't necessarily mean that results are coming out more competent, but it would at least mean that residency training has improved" (P13). Participants explained that these types of short-term questions are more relevant to the program implementation stage than asking long-term outcome questions, such as, "did [the transition to CBD] make any difference to the quality of the physician? I don't know" (P14) or "what is the impact of CBME? [It requires] a bit of improvising the program evaluation that we have" (P15). Most participants acknowledged that these type of outcome or impact questions may not be answerable for many years, are influenced by various factors, and are more appropriate to higher RCPSC level questioning.

In summary, participants discussed how asking meaningful questions that are relevant to the program implementation stage can help their CBD programs build their capacities to use evaluation findings.

**Establish Buy-In for Program Evaluation to Increase the Likeliness of Using the Findings.** Participants suggested that CBD programs can build their capacities to use evaluation findings by establishing buy-in for program evaluation, which increases the likelihood that findings will be used. Participants noted that one way to establish buy-in can involve a level of participation in the evaluation that leads to a greater chance that the findings will be applied. For example, when a faculty member or resident, "see [that] 'I actually gave some feedback on this part of the program that needed some improvement,' four months later, 'there's some actual improvement there,'" (P8) they will buy into the program evaluation process. When buy-in develops, "that's when you get people who are . . . much more engaged in evaluation, and they see some value in that their opinion is valued. And some change may actually occur," (P8) and they are more likely to use the findings themselves and participate in the evaluation process again in the future. As participants are "learning more about program evaluation, [we] . . . expect [there] to be some shift in how [we] see the value of the change as well" (P6). In learning by doing program evaluation, participants understand the value in using findings to create evidence-based change.

Participants also indicated that when the data is used to inform decision making, buy-in increases because they feel that their voices are heard and validated. They explained that this cycle reflects how the use of findings establishes stronger buy-in, which in turn strengthens commitment to using findings. Participants noted that "we get program evaluation findings. We have to discuss them and decide on a change. Change our idea or not and then implement it" (P15). They shared how "we do a reasonable job of it, but I think we could do so much better and have more data to be able to defend any decision" (P1) and "to defend certain things. And also then to justify [our] requests as well, when [we] asked for things" (P9). Participants noted that buy-in is solidified after seeing how feedback has informed changes to curriculum content delivery and has increased further use of findings. For example, participants described how they "have basically a new rotation and a new clinic. We are inventing a clinic basically. Or sometimes it's a new teaching module" (P15) due to feedback from clinical supervisors about a challenging EPA to observe. To establish a use of findings cycle, participants "made a

commitment when we made all these changes that three years out, we would look at that again . . . because we know it takes a few years to really get things going" (P1). Participants described that when faculty bought into and made a commitment to review the change in the future, it increased the use of findings because faculty were aware of follow-up. Furthermore, participants noted that establishing buy-in to program evaluation can increase the likeliness of the findings being used to make change at the national level. For example, participants indicated that they were providing feedback at the national level to, "the Royal College requirements and be able to change the curriculum that way and identify which [EPAs] need to be amended" (P7).

In summary, participants discussed how establishing buy-in to program evaluation to increase the likeliness that findings will be used can help their CBD programs build their capacities to use evaluation findings.

**Share Findings with Faculty, Residents, and Other Programs in Ways That Encourage Use.** Participants suggested that CBD programs can build their capacities to use evaluation findings by sharing findings with faculty, residents, and other programs. They noted that when faculty and residents are involved in the evaluation process, it supports their commitment to learning about the findings and enacts ownership of the findings. Participants advocated for acting "transparently . . . [as] this 27-page report . . . was shared with basically everyone who engaged in the process which would also be our stakeholders that would be impacted" (P1). Specifically, "it's important, especially for something of that scope to be transparent about things so that way, whether it's good or bad you agree or disagree, at least, everyone knows and has access to information" (P1). Participants noted that sharing the findings with everyone involved in the evaluation built an awareness and understanding about the process of evaluation and established value for what was evaluated. In one example, findings were accessed by "the faculty on the ground in terms of structuring the training [and they were] advised of changes in requirements, expectations, and/or structure" (P2). By sharing findings from an evaluation in which faculty and residents participated, faculty and residents felt ownership to brainstorm solutions and to make the suggested changes.

Participants also emphasized the "continuous quality improvement on the training of the residents. So, whenever a review is done, then, we do actionable things to that module to tweak it . . . And then, we go through the cycle again" (P2). In other words, sharing findings and following up in a review cycle supports the use of findings because it holds faculty and residents

accountable for changes and there are future, scheduled check-ins. In addition to this internal, specialty/subspecialty review cycle, participants noted that they "report to the Department of Health and Wellness from a faculty perspective to see if faculty are participating in the CBD implementation . . . So reporting to PGME and reporting to Health and Wellness" (P13). Reporting findings to PGME and external departments encourages ownership of acting on the findings because future check-ins from these departments may occur. Finally, some participants noted that they "do share . . . it's more [some have] used more resources from other universities than the reverse because [they] launched at the very end of the process" (P15). Participants indicated that sharing evaluation findings with other specialties/subspecialties can encourage the use of findings because those groups may inquire about updates or changes to the implementation process as a result of the findings.

In summary, participants discussed how sharing findings with faculty, residents, and other programs in ways that encourage follow-up cycles and ownership can help their CBD programs build their capacities to use evaluation findings.

**Understand Evaluation as Helpful for the Accreditation Process so Program Directors Are More Likely to Use the Findings.** Participants suggested that CBD programs can build their capacities to use evaluation findings by viewing evaluation as advantageous for the accreditation process. Participants noted that "getting involved in more formal program evaluation really was kind of spurred on by accreditation, but I think a nice offshoot of it was that it showed us how we could develop a formula maybe for ongoing program evaluation" (P12). As such, they "did use . . . that feedback [to learn more about] what's actually happening . . . [but] it's mainly for accreditation" (P9). Additionally, participants indicated that "the Royal College is looking to see how we're evaluating certain things and how we're making those changes things like that, so . . . [program evaluation] would be a nice feather in your cap at the very least" (P7). They also described how using evaluation findings can "show that you're kind of evaluating the program on an ongoing basis and implementing changes" (P7). Using findings to implement changes is "something that [we are] sure that [the RCPSC accreditors would] be looking into, so it's important to have something along those lines in our reports" (P7) and it can make program directors more likely to use the findings. Participants confirmed that "the main focus of the Royal College for accreditation now is a focus on continual improvement" (P3). Therefore, "being able to have data to then be able to reflect upon and change would meet that

need for continual improvement and so that's what they're trying to see is that programs have that ability" (P3). Participants added that "accreditation . . . is to evaluate your program that you are providing appropriate education . . . [and] that's like the big kahuna of program evaluation when the Royal College says here comes our evaluators" (P8) because "as part of the Royal College review . . . there is always a portion that asks about, 'How do you evaluate the quality of your educational program?'" (P10). Participants emphasized that program evaluation "was largely driven by accreditation. So, we underwent accreditation at [university name] and as preparation for that, our PGME office had suggested that we have this [program evaluation] for our program as a whole" (P12).

Participants also indicated that they were more likely to use findings when they viewed evaluation as helpful in preparation for the accreditation process, and also as valuable after the accreditation process. During accreditation, "you really look at your program with a fine-tooth comb, and through a microscope. And that, in [and] of itself, that really is what the reviews are about is that you reflect on your program" (P8). Participants acknowledged that, following accreditation feedback, "no program has every single thing—not one of them . . . but, then you look at remedying that accreditation [via] program evaluation" (P8). Similarly, participants commented that "for sure the accreditation process is a big part of it, but accreditation is there for a reason, honestly, for the programs to continuously improve" (P5). Participants noted that using the findings from an accreditation review resembled using the findings from program evaluation. However, participants explained that accreditation offers an external, fresh perspective whereas program evaluation offers an in-depth, ongoing reflection from faculty—both findings are helpful and complementary.

In summary, participants discussed how understanding evaluation as helpful for the accreditation process can help their CBD programs build their capacities to use evaluation findings.

**Summary for Phase 2**

Of the 36 program directors who were invited to participate in Phase 2 of this study, 15 consented to be interviewed about the ways in which they or others in their CBD programs are engaging or not engaging in program evaluation and how they or others in their CBD program can build their capacities to do program evaluation and use evaluation findings. Table 14 at the

beginning of this chapter, provides a summary of the Phase 2 theme findings from the program director interviews.

Participants whose programs are engaging in program evaluation discussed how their programs are: (a) struggling to engage in program evaluation because of limited available resources and buy-in; (b) using ad hoc evaluation methods; and (c) using a team-based evaluation format.

Participants also discussed ways that CBD programs that are beginning to or are not engaging in program evaluation can build their capacities to do program evaluation. Specifically, they can: (a) develop expertise in their program evaluation abilities; (b) acquire program evaluation resources for guidance; and (c) advocate for clear program evaluation expectations from the RCPSC, the PGME office, and within their specialties.

Finally, participants discussed ways that their CBD programs can build their capacities to use evaluation findings. Specifically, programs can: (a) ask meaningful questions that are appropriate to the stage of the program; (b) establish buy-in for program evaluation to increase the likeliness of using the findings; (c) share findings with faculty, residents, and other programs in ways that encourage use; and (d) understand evaluation as helpful for the accreditation process so program directors are more likely to use the findings.

These Phase 2 findings informed the Phase 3 interview guide to support answering the Phase 3 questions. Specifically, in Phase 2, program directors indicated that they are struggling to engage in program evaluation because of limited available resources. In Phase 3, my interview questions asked the program evaluators about available program evaluation resource suggestions and applicable evaluation methods. Additionally, in Phase 3, my interview guide elaborated on how program evaluators can help program directors to develop expertise in their program evaluation abilities. Finally, in Phase 3, my interview guide prompted program evaluators to suggest how to establish buy-in for program evaluation to increase the likeliness of using the findings.

# Chapter 6: Phase 3

In this chapter, I describe the methods and results of Phase 3, during which I investigated program evaluators' perspectives about how they are currently supporting CBD programs in program evaluation and how they can help CBD programs build their capacities to do program evaluation and use evaluation findings. For this component of the study, I interviewed a selected group of program evaluators in medical education. The purpose of this phase was to explore the following research questions:

- Question 5: How are program evaluators currently supporting CBD programs in program evaluation?
- Question 6: How can program evaluators help CBD programs build their capacities to do program evaluation and use evaluation findings?

## Sample

I used purposeful sampling to recruit program evaluators with the requisite knowledge (i.e., evaluating medical education programs) to be interviewed for this study (Patton, 2015a). Eligible participants included Canadian English-speaking members of the AEA. Based on the relevance of their designated affiliations, I included members from two different AEA TIGs: Health Evaluation and Health Professions Education, Evaluation and Research. I obtained the eligible evaluators' email addresses from the AEA TIG website. Based on the AEA website member directory, 33 eligible evaluators received the invitation for an interview. In addition to these 33 evaluators, I contacted 4 program evaluators who evaluate medical education programs in Canada but are not AEA TIG members. I also contacted 15 program evaluators associated with Canadian English PGME programs. In total, I contacted 52 program evaluators, and 12 consented to be interviewed for Phase 3. Appendix M shows the sampling process and the participants who agreed to be interviewed.

## Instrument Development

Based on the literature review presented in Chapter 2 and the findings from Phase 1 and Phase 2, I developed the semi-structured interview guide for Phase 3 (see Appendix P) for participants who evaluate CBD medical education programs. Table 14 presents the research questions that guided Phase 3 and the 13 corresponding interview guide questions. The guide begins with an introductory script and includes open-ended questions with additional probes to allow for expansion and idea prompting. These prompts offered insight into program evaluators'

thoughts on how program evaluators are currently supporting CBD programs through program evaluation, and how program evaluators can help CBD programs build their capacities to do program evaluation and use evaluation findings. I piloted the guide with my supervisor and another program evaluation researcher to ensure it was clearly worded and included relevant questions or probes (Seidman, 2019).

**Table 15**

*Specifications for Phase 3 Interview Guide: Program Evaluators of Medical Education*

| Research questions | Corresponding interview questions |
|---|---|
| 5. How are program evaluators currently supporting CBD programs in program evaluation? | Question 1<br>Question 2<br>Question 3<br>Question 4<br>Question 5 |
| 6a. How can program evaluators help CBD programs build their capacities to do program evaluation? | Question 6<br>Question 7<br>Question 8<br>Question 9<br>Question 10 |
| 6b. How can program evaluators help CBD programs build their capacities to use evaluation findings? | Question 11<br>Question 12<br>Question 13 |

*Note*. CBD = Competence by Design.

**Data Collection Procedures**

I emailed the eligible 52 program evaluators a letter of information and a consent form (see Appendix N and Appendix O) inviting them to participate in an interview for Phase 3. The information letter asked the program evaluators to reply to me if they were interested in participating. Each interview was scheduled at a time that was convenient for the interviewee, took place using Zoom (https://zoom.us), and lasted approximately one hour. Prior to the interview, all participants signed and returned an informed consent form to me. I recorded and transcribed each interview verbatim. Each transcription was sent to the corresponding interviewee for approval and feedback. I applied any changes identified by the interviewees. I gave each participant a $10 Amazon gift card to thank them for participating in the interview.

**Data Analysis**

Similar to Phase 2, I followed Miles et al.'s (2019) systematic step-by-step approach to acquire meaning from data and confirm findings. First, my supervisor and I analyzed the data by tracking emerging themes from the interviews, beginning with deductive codes based on my literature review. We used NVivo software to organize the data and record our interpretive process (Horsley et al., 2016). Per Miles et al. (2019), we began with "an a priori list of researcher-generated codes" (p. 69) based on the interview question guides about context, stakeholders, evaluation design, data collection, data analysis, dissemination of evaluation findings, use of evaluation findings, reasons for engaging in program evaluation, confidence in abilities, training received, resources, impact of access to training and resource availability, program needs to build capacity to do program evaluation, and possible suggestions for changes. We refined the coding system while I was initially transcribing the interviews, and then we read through each transcript for a second time and included inductive codes to generate a comprehensive and reflective code list. We then read each transcript for a third time and applied themes by reflecting on our a priori list of codes and highlighting data to exemplify or refute a priori codes. We reviewed each transcript for a fourth time and revised our codes as an inductive exercise where "other codes [not on the a priori list] emerge progressively during data collection" (Miles et al., 2019, p. 74) or, alternatively, "some codes do not work. This issue calls for doing away with the code or changing its type" (Miles et al., 2019, p. 75). Next, we reviewed the revised codes noted on each transcript to update the a priori list, group the codes into overall themes, and reduce the number of themes, which is also known as "second cycle coding" (Miles et al., 2019, p. 79; see Appendix Q for the final Phase 3 NVivo codes). Then, we used the themes, codes, and coded quotations to create a data analysis matrix, which we used to highlight and understand the relationships between themes and codes, and to make note of irregular or contradictory findings (see Appendix R for a sample matrix). We leveraged data from this matrix to draft initial findings from the Phase 3 data and share explanations for the code relationships and themes. I invited each interviewee to review the preliminary study findings and offer their perspectives (Creswell, 2014). I applied their feedback to revise the transcripts, theme codes, and matrix. Although I conducted interviews using Zoom (https://zoom.us), I did not analyze participants' non-verbal cues. I only analyzed the wording in the recorded interviews that I

transcribed. Each interviewee was identified using a pseudonym made up of the letter "E" for program evaluator and a number based on the interview order (e.g., E1, E2, etc.).

Finally, in my qualitative analysis, I applied Patton's (2015a) suggestions to avoid quantifying my qualitative data. Specifically, I intentionally decided to not quantify or count the number of participants who mentioned ideas that I associated with representing each theme. In this way, I followed Patton's (2015a) arguments to remain qualitative because "it's not how many said something that matters" (p. 557). Therefore, I have omitted the number of respondents who shared each theme, so I can emphasize the participants' message content and not take away from their perspectives. Instead of reporting all quotations collected from participants, I have provided exemplar quotations to establish a "verbal display that represents and presents data vividly about the study's phenomenon of interest" (Miles et al., 2019, p. 324). I have selected these example quotes because they accurately highlight the interviewee perspectives about engaging in evaluation of a CBD program.

***Trustworthiness***

Similar to Phase 2, I used the following three criteria to assess the trustworthiness of my qualitative research and to ensure that the findings validated the experiences of the individuals under investigation: (a) credibility, (b) confirmability, and (c) transferability. I used credibility to evaluate trustworthiness by member-checking (i.e., participant verification of the data's accuracy in representing their perceptions and experiences; Lincoln & Guba, 1985). After coding the qualitative data using NVivo, I verified the main ideas with the interviewees to ensure accuracy. I used confirmability to ensure trustworthiness by comparing the interviewee main ideas in NVivo to find similar or diverging data to identify meaning from the study. These similarities and differences inform my discussion in Chapter 7. Finally, I used transferability to evaluate trustworthiness by comparing whether similar findings have been published in journals about research in program evaluation of CBD programs. Furthermore, I used notetaking to document my assumptions, questions, and decisions during the data interpretation. This descriptive notetaking strengthens the trustworthiness of the findings by making this study replicable in other settings. The findings of Phase 3 are summarized and presented in monograph form in the following section.

## Findings

### *Characteristics of Participants*

Of the 52 program evaluators who I contacted to participate in an interview, 12 consented to be interviewed. The interviewees included program evaluators associated with the AEA's Health Evaluation and Health Professions Education, Evaluation and Research TIGs, program evaluators associated with Canadian English PGME programs, and program evaluators who were known to evaluate medical education programs in Canada but are not listed on the AEA website.

**Table 16**

*Phase 3: Theme Findings from Program Evaluator Interviews*

| Research Question | Theme Findings |
|---|---|
| ***5: How are program evaluators currently supporting CBD programs in program evaluation?*** | (a) Responding in a reactive way as a temporary and external evaluation consultant<br>(b) Collaborating with interested stakeholders as internal evaluation experts from a university<br>(c) Leading an evaluation approach that is manageable for program directors with limited time<br>(d) Using methods that complement the PGME environment<br>(e) Using evaluation processes that lead to the publication of findings<br>(f) Offering professional development sessions about program evaluation to support accreditation activities |
| ***6a: How can program evaluators help CBD programs build their capacities to do program evaluation?*** | (a) Using a participatory evaluation approach<br>(b) Offering evaluation resources and expertise as guidance<br>(c) Leveraging existing data<br>(d) Advocating for clearer program evaluation policy expectations from the RCPSC |
| ***6b: How can program evaluators help CBD programs build their capacities to use evaluation findings?*** | (a) Recommending that program evaluation is included at the beginning of program design and implementation<br>(b) Including the stakeholders who will be using the findings in the evaluation itself<br>(c) Focusing stakeholder questions on meaningful program areas over which they hold authority to make changes<br>(d) Encouraging the use of program evaluation approaches that are appropriate to the CBD implementation context<br>(e) Interpreting the findings alongside the stakeholders and co-develop a manageable follow-up plan<br>(f) Encouraging programs to share findings, which establishes an accountability cycle to make changes based on findings |

### *Research Question 5: How Are Program Evaluators Currently Supporting CBD Programs in Program Evaluation?*

In the sections that follow, I describe how the participants are currently or could be supporting CBD programs in program evaluation. Specifically, program evaluators are or could be: (a) responding in a reactive way as a temporary and external evaluation consultant; (b) collaborating with interested stakeholders as internal evaluation experts from a 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 a Temporary and External Evaluation Consultant.** Participants indicated that they are currently or could be supporting CBD programs in program evaluation in a way that is more of an afterthought, that is, as temporary or external evaluation consultants. Program evaluators explained that in their "consultant role, it's much more of 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). Participants also noted that a reactive evaluation response "is the way that we don't favour . . . when they become interested in program evaluation after the fact" (E2). That is, CBD programs have "already been developed, it's already been implemented. And then they sort of want to collect some data on how things are going after the fact" (E2). One problem with being an external evaluator who is invited in after program design and implementation is that the evaluation can be seen as a formality, and "all the work and resources get put in and then a report gets shoved and hidden in a drawer and collects dust" (E2). Participants noted that one of the reasons there are temporary and external evaluators is because the RCPSC is "not saying we want all you programs to do program evaluation, and here are some resources and supports for everybody to do it" (E2). Program evaluators explained that "it absolutely is up to the individual program directors to take on the responsibility of program evaluation. And I would even say to even decide if they're going to do any program evaluation" (E2). Although program evaluators indicated that the early stage of program development "is prime opportunity to integrate program evaluations," they also indicated that program directors "are so overwhelmed by all the changes and everything that they need to do just in their day-to-day that the thought of adding and integrating something else within it really

overwhelms their mind" (E2). The dependency on overloaded program directors to take the initiative with program evaluation explains why program evaluators indicated that they have "not come across an RFP [request for proposal] for that. So that's why [many evaluators are] not doing it because nobody has invited [us] to do it" (E1).

In summary, participants discussed how, when invited, they are currently or could be supporting CBD programs in program evaluation by serving in a reactive way as temporary or external evaluation consultants.

**Collaborating with Interested Stakeholders as Internal Evaluation Experts from a University.** Participants indicated that they are currently or could be supporting CBD programs in program evaluation by collaborating with interested stakeholders through a university's internal evaluation team. Program evaluators stated that if they "feel like [program directors are] not really committed to evaluation, then it's not something [we would] want to get engaged in" (E1). When program directors are interested and motivated, participants indicated that they "have been engaged in program evaluation with Competence by Design [that] has been much more about a collaborative fashion" (E2). Specifically, participants' responsibilities were to "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). Interested stakeholders included "their program director, their program administrator and their residents for a focus group and surveyed faculty and residents" (E10) as well as "the hospital supervisors, academic advisors and competence committee members, [and] the clinical supervisors" (E8). Participants mentioned interest from residents and described one "resident [who] was a superstar who was . . . pushing and leading [the evaluation] and [she] and [her program director] were going back and forth" (E5). Program evaluators highlighted stakeholder involvement as necessary because those are the experts in existing data sources and in data sources that need to be created to answer the evaluation questions. For example, "somebody says how about a patient survey? Okay, do we have any existing patient surveys? No. Okay. We need to develop one" (E5).

Participants also discussed their experiences working in an internal position within a university's evaluation unit. For example, participants indicated that program directors would "reach out to us telling us we need to do this and that and we think of evaluation framework and how we're going to approach the concerns. And then we'll provide guidance along the way"

(E4). Other program evaluators contextualized their work as being within "an internal evaluation unit . . . in medical education . . . in the undergraduate environment, [because] it's a lot more uniform . . . it's just one learning environment survey for everybody" (E3) and they contributed to "a multidisciplinary group . . . responsible for providing program evaluation" (E6). Thus, despite a participant's position within an internal PGME evaluation unit, they did not "know how much [post graduate CBD evaluation] exactly happened . . . we weren't doing anything" (E6) because their focus was on undergraduate medical education. Additionally, some participants were "also hearing . . . the Royal College as one source . . . were trying to provide some support to do evaluation or monitoring or gathering data around [PGME] . . . so, [we weren't] feeling quite so guilty" (E6).

Moreover, participants noted that "where they found our [evaluation] services really, really useful is that we're able to kind of get into the smaller programs without the risk of kind of the anonymity or confidentiality of the residents" (E3). Participants suggested that smaller postgraduate programs doing program evaluation can collect honest, constructive feedback from residents and faculty to encourage responses without fear of retribution by leveraging an evaluation unit. Program evaluators described that "for [smaller programs], they're finding the value to having an internal/arm's length internal unit to do this work" (E3). However, "the idea that we were just waiting for them to come and seek help. I don't know if this is helping or not. Because we're not getting a lot of requests" (E4). Thus, program evaluators identified a disconnect despite "the Royal College supporting program evaluation to a great extent [at] . . . summits and many presentations about program evaluation . . . [about] where to get help on program evaluation" (E4). Participants highlighted this lack of clear communication as a problem because while program evaluators were waiting to be asked, program directors who were interested in evaluation were unsure who to ask about it, were occupied with other concerns, or were unaware of the existence of program evaluation support.

Furthermore, participants noted that their positions are within "an internal unit that services the entire undergraduate and postgraduate education and as well as health professions education. So, we are of course limited in resources" (E3). They also indicated that it is "senior leadership . . . [who] directs and gives priorities [to] program areas that we should be working in" (E3). Participants stated that, to ensure fairness, senior leadership does not "want to use our services for some programs and not others, [thus they are] still in the process of figuring out

[how] to support the entire residency programs and in a systematic way" (E3). Program evaluators indicated that "postgrad, because it is so decentralized and all the programs do their own things, [programs] are all at very different stages of implementing CBD" (E3), which complicates how senior leadership directed their internal evaluation unit. Participants also stated that they have "tried . . . doing evaluation work that will meet the common good or broad needs to inform planning and decision making for in postgrad generally" (E6), but "the postgrad deans are in a place where they're trying to figure out exactly how to use us to get the most bang for their buck" (E3). Finally, program evaluators within internal units questioned evaluating CBD using "an overall common framework for evaluation—would that meet needs? Or would you have to come up with tailored evaluation strategies for each of the programs?" (E6).

In summary, participants discussed how they are currently or could be supporting CBD programs in program evaluation by collaborating with interested stakeholders while working in an internal position as part of a university's evaluation unit.

**Leading an Evaluation Approach That is Manageable for Program Directors with Limited Time.** Participants indicated that they are currently or could be supporting CBD programs in program evaluation by leading an evaluation approach that is appropriate for program directors with limited time. For example, some program evaluators led a "rapid cycle evaluation when they transitioned [to CBD] and that informed a lot of the adjustments that they made" (E8). Participants indicated that this approach was possible because the CBD faculty lead "understood the value of program evaluation and advocated for funds to be allocated to that" (E8). However, "it wasn't really rapid. In an ideal world, it would get done in the first six months of implementation and adaptations would be made immediately . . . we should have been evaluated within the first six months" (E8). In this particular case, the program evaluation used the Core Components Framework and "contracted [a research team] to do the data collection in terms of documenting their intention, and then conducting interviews and focus groups with the stakeholders, and then drafting a technical report" (E8). Other participants described the use of "a continual improvement process, or developmental evaluation process" (E7). The ongoing, flexible nature of developmental evaluation allowed for stakeholders to make "changes as we've gone because we realize it's not perfect" (E7) and to "think about a seamless way of incorporating it into what's already being done" (E7). Finally, some participants led a "realist evaluation of the implementation of CBD" within a Canadian university's PGME department

(E10). A realist evaluation is "a complexity-based evaluation plan, and it's not a logic model. It's not an evaluation framework" (E10). Participants explained that the realist evaluation's objective was to explore "the questions of, 'What works? For whom? In what circumstances? And, why?'" (E10). Moreover, realist evaluation is "not a utilization focused evaluation design where you . . . measure this thing over time, and then we'll change this and see if it results in improvement or not improvement" (E10). Instead, implementation is contextually bound to learn more about "what's happening, and when things are working, why are they working, but it's also a complexity theory, which means that just because it works in anesthesia doesn't mean it's going to work in internal medicine" (E10).

Participants who are leading evaluation approaches explained how they made the approach achievable for program directors with limited time, highlighting that stakeholders "have to still be bought into it enough. So, it's this fine balance, getting their buy-in, but not requiring too much" (E7). Program evaluators who used the rapid cycle evaluation explained that program directors "completed their program evaluation matrix . . . and then we might have communication through email, but they don't have to meet with [program evaluators] again until the technical report was being reviewed" (E8). Additionally, "all the data collection was paid for under this [funding] allocation from the post grad office, so [program directors] didn't have to worry about facilitating focus groups or conducting interviews or analyzing the data which was huge" (E8). Program evaluators indicated that the time commitment required to analyze data is where program evaluation often gets stuck because program directors "don't have the time to do that" (E8) or "the time to think about evaluation" (E4). Participants also emphasized that "these program directors don't have a lot of time . . . [and] at the program director level, there has to be protected time . . . to meet" (E5). To make program evaluation manageable, program evaluators suggested prompting for "short-term and maybe intermediate outcomes" (E5). Participants also stated that "the barriers for people doing evaluation are pretty much the same for medical education programs as other programs . . . lack of capacity, lack of time" (E1). They noted that "it really comes down to resources and availability of time. Everybody wants to do a good job, but we only have so many working hours in the day" (E3). However, they insisted that "they're not barriers on [the program evaluator] side, because [although we] can't give them more time, [we] can efficiently develop something so they don't have to devote as much as their time to it" (E1). Finally, participants explained that "it's not going to take them as much time if they're

working with a professional evaluator. And they're going to get a really high-quality piece of work. So, the barriers aren't on the evaluator side" (E1).

In summary, participants discussed how they are currently or could be supporting CBD programs in program evaluation by leading evaluation approaches that are manageable for program directors with limited time.

**Using Methods That Complement the PGME Environment.** Participants indicated that they are currently or could be supporting CBD programs in program evaluation by leveraging methods that are suited to the overloaded PGME environment without requiring additional resources. For example, some participants explained that their data collection involved "qualitative data purely" during "interviews with different people" (E4). With the use of interviews and focus groups, participants stated that the data collection "was all qualitative really. [Program evaluators] focus on the experiences of the stakeholders" (E8), and other program evaluators leveraged "interviews, focus groups and surveys" (E10). To ensure consistency during data collection, participants noted their "set interview and focus group protocols. So, it's a pretty standardized process now . . . and we explore the stakeholders' experience . . . so the reality of what has gone on" (E8). Some participants led a group who gathered and analyzed data, and met weekly, that included "four [people] . . . and that didn't include the research assistants . . . [and] included one . . . project manager [who] oversees . . . and does do some data collection" (E8). As such, participants insisted that successful program evaluation data collection requires a dedicated personnel position. Finally, participants explained how they designed "the interview guide around exploring [program directors' and residents'] experiences of implementing what went well, what didn't go well?" to learn more about "why something worked or didn't work?" (E10).

Regarding quantitative data collection, participants emphasized that the focus of evaluation support was "more of the methodological tools, designing the survey, refining the survey" and reviewing "the response rate" (E5). Participants specified that to form a consensus, their evaluation "was a modified Delphi . . . [using] a survey [of] program directors" (E5). Some participants also shared that although "we do have ethics to look at the quantitative data because we've got the biggest data set Canadian wide in terms of EPA assessments . . . we didn't really use any quantitative data. It was all qualitative really" (E8). Other program evaluators included "a framework and the one survey that was distributed to the students, some self-reflection . . .

[with] more data sources and data indicators that we are going to use, but we haven't actually used them yet" (E4). Moreover, 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" (E4). Participants remained hopeful that "maybe after some time, we can have more assessment data and we can perform the analysis" (E4).

Participants also shared that "one of the challenges was faculty burnout . . . [because] they're already busy enough and asking them to check a reflection, that's not part of their course, it's part of the program evaluation—that didn't go over well" (E7). They emphasized that "one of the challenges [with data collection methods] is faculty responsibility and adding to their plates. Because they're already overburdened" (E7). Despite the efforts to use either qualitative, quantitative, or mixed methods in a hectic PGME environment, "if they do decide to do program evaluation, the big problem is that they would have to commit to doing any of this [program evaluation] within the same resources and budgets that they have for their program" (E2). Participants explained that it's a "resource of bandwidth thing" and "there were other projects requirements in the post grad environment" (E6). Participants concurred that the reason they "haven't gotten into evaluation in our postgraduate programs [is because of], unfortunately, setting priorities within given resources that we have" (E6). Thus, balancing existing responsibilities while transitioning to CBD has prevented program evaluation, which "just hasn't been occurring" (E2). The implementation of CBD has left program directors "so swamped in the transition" and "their time is divided between clinic and between other educational duties, and they don't have time to think about improving the program and looking for ways to do program evaluation" (E4).

When program evaluation is postponed, program evaluators explained that it is "because of many potential issues again, time, money. Because they would have then needed to again either hire a consultant or something to help them actually carry out the program evaluation" (E2). The PGME program must consider where it's "going to allocate money . . . [because] you put something in one place, you got to often take it away from something else" (E8). Participants noted that at an RCPSC evaluation summit, "there was this question about sustainability of the work" (E4). Across Canada, the program evaluation PGME typically resembled "one person or two people in a unit providing program evaluation support and they have problems with funding their work as well. And you have to apply for different grants to keep the work going" (E4) and

"the biggest challenge is always funding" (E5). When program evaluation was considered a priority by administration, "they found the money, they're like, yes, we need this to happen. We're going to find the money" (E5). Program evaluators emphasized that priorities of value (i.e., program evaluation) require funding to maintain their initiatives.

In summary, participants discussed how they are currently or could be supporting CBD programs in program evaluation by using methods that complement the busy PGME environment without requiring additional resources.

**Using Evaluation Processes That Lead to the Publication of Findings.** Participants indicated that they are currently or could be supporting CBD programs in program evaluation by supporting program directors in applying evaluation processes that lead to the publication of findings. For example, participants 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 "had a standardized process and really follow[ed] the Core Components Framework" (E8) to write a report, share findings, and replicate a similar evaluation with other program specialties. Program evaluators specified their role "to help develop an evaluation matrix [and] to help implement the evaluation process up to the point of reporting" (E2). The application of an evaluation framework supported program directors who were "interested in program evaluation . . . [who] want to share with other programs across the country or even internationally" (E2). Participants explained how they guided the process to prepare program directors for sharing "some aspect of some program whether it's evaluating a new initiative, whether it's evaluating the effectiveness of an initiative, or evaluating a program . . . one is evaluating CBD in postgrad" (E10). Participants noted that the structured evaluation findings were shared by publishing them in academic journals.

Participants also noted that program directors are "enticed by publication . . . because Competence by Design is so new . . . and they really want to publish it" (E2). Participants indicated that they have been "approached to help develop a small evaluation project and of course, publish the findings" (E2). Some participants shared how they had "eight programs in process and our intention was always to support them in the technical reports converting those to publications. So, that was the carrot for the program leaders to be involved" (E8). Participants concurred that inclusion in the publication was one reason why program directors were enticed into program evaluation activities. For example, some program directors had "come up with the

idea of designing the small project and figuring out how they [could] collect some data . . . and then publish it for the purposes of sharing and figuring out what they [could] do next or what they [could] do differently" (E2).

In summary, participants discussed how they are currently or could be supporting CBD programs in program evaluation by using evaluation processes that lead to the publication of findings.

**Offering Professional Development Sessions About Program Evaluation to Support Accreditation Activities.** Participants indicated that they are currently or could be supporting CBD programs in program evaluation by providing program evaluation professional development to support accreditation. For example, program evaluators noted that "evaluation experience and training has taken place more so within the realm of professional development and within the work setting" (E2). During the curriculum mapping, program directors "identified where the courses that [residents will] take will align with various competencies . . . and had a series of workshops with faculty to align the courses to the competencies" (E7). Participants also indicated that program evaluators "teach evaluation" at postsecondary locations and do "a lot of workshops for different health professions" (E1). Furthermore, they noted that stakeholders from "different universities [are] talking about program evaluation in medical education [at] the Royal College program evaluation summit presentations" (E4). Finally, participants noted that the "people [who] are really interested in program evaluation are the ones who attend, [but it's] usually not [all program directors]" (E4).

Participants also shared that interest in program evaluation is largely tied to accreditation and acknowledged that "medical education programs require evaluation as part of accreditation. So, that is one of their criteria to get accredited—doing ongoing program evaluation" (E1). Program evaluators' responsibilities with "evaluation in medical programs . . . [are to support] faculty. . . [as] the lead evaluator working towards accreditation" (E7) and to focus on "new work that we needed to get embedded and demonstrate that were happening to meet continued accreditation requirements" (E6). For example, participants' internal evaluation units were "heavily involved in the evaluation of the curriculum renewal for the medical school because that was going to be demanded to ensure they maintained accreditation standing. So, that was a must do" (E6). Furthermore, participants explained that they have "had some really more high-profile, high-risk, significant drivers delivering some streams of evaluation and solid evaluation

reporting for them that were required by accreditation. So, it was kind of non-negotiable" (E6). Therefore, despite an interest in learning more about various aspects of specialty/subspecialty programs "there were other priorities that we had to address before that for accreditation purposes" (E3). This means that "all quality assurance is accreditation driven. They happen on a cycle and in between their accreditation times they work with the programs that have decided to kind of fix itself until the next accreditation" (E3). As such, "when we hear about accreditation for post grad is coming, we're going to get knocks on the door" (E3). In between accreditation cycles, there is an "internal site visit, basically like accreditation surveys for each program . . . and that in itself is kind of program evaluation" (E3).

In summary, participants discussed how they are currently or could be supporting CBD programs in program evaluation by offering professional development sessions about program evaluation to support accreditation.

***Research Question 6a: How Can Program Evaluators Help CBD Programs Build Their Capacities to Do Program Evaluation?***

In the sections that follow, I describe how the participants can help CBD programs build their capacities to do program evaluation. Specifically, program 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.** Participants indicated that they can help CBD 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. Program evaluators explained that "every time I do an evaluation, I want people that I'm working with to learn from the process so that they might pick it up and see the value of it and it's not something scary" (E5). Engaging stakeholders in the process to learn by doing offers an opportunity for "evaluation capacity building" (E5). Participants indicated that they "really want it to be a learning experience" for residents and other stakeholders "to self-select into" the evaluation to support their journey to "be a researcher . . . work at a tertiary hospital . . . [or] go on to be academic researchers and evaluators" (E5). Moreover, participants noted that active stakeholder participation in evaluations developed knowledge of program evaluation "because they're trying to evaluate towards all of the standards that they have to meet and trying to figure

out exactly where the gaps are and where the challenges are, where they're doing well" (E3). Participants indicated that one reason for applying a participatory approach is because "there's no way we can provide support to 60 plus programs, and we are only two people [on the program evaluation PGME team]" (E4). Therefore, "the only way there is to build the capacity of programs to do their own evaluations . . . [and] they can learn when they start doing evaluations" (E4). Finally, participants noted that engaging stakeholders in program evaluation leadership initiatives helped the stakeholders develop an evaluation mindset.

When using a participatory approach, program evaluators emphasized *prompting evaluation* tasks up into manageable, specific, and achievable chunks. When supporting "physicians in evaluation, they start at ground zero. They'll come to me and say, 'I want to evaluate this program, but I know nothing about evaluation'" (E5). When planning with the end in mind, participants indicated that they prompt program directors to "talk about [their] main goals," focus on "what are the objectives," and "work [their] way back" using "evaluative language" (E5). Some participants leveraged this learning opportunity to co-create and scaffold an evaluation matrix to generate appropriate and accessible program evaluation questions. They shared that when "going through the process of creating the evaluation matrix and thinking about the questions and data sources, [the program director] said that he is learning along through the process" (E4). Program evaluators also emphasized the need to focus the evaluation design and to specify the area under investigation, which was essential for referring back to "walk along the evaluation matrix. So, then designing it for them and they have it in front of them" (E5). Although some program directors "are just way more adventurous, or they have faculty members who are in the forefront of thinking about CBD" (E3), others need smaller steps and guidance to structure their learning about program evaluation. Participants who scaffolded program evaluation into smaller steps allowed program directors to "try little things . . . instead of changing the entire curriculum all at once and then evaluating little pieces of it" (E3).

In summary, participants discussed how they can help CBD programs build their capacities to do program evaluation by applying a participatory approach to encourage small steps towards learning by doing.

**Offer Evaluation Resources and Expertise as Guidance.** Participants indicated that they can help CBD programs build their capacities to do program evaluation by providing evaluation resources and expertise as guidance in order to maximize program directors' time

concerning evaluation. 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, participants 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*. Program evaluators emphasized the provision of simplified evaluation resource templates as "a roadmap" (E3) towards establishing a culture of evaluation and to save time. Participants advocated for "some broader, maybe cross cutting kinds of tools or whether it's frameworks and supporting tools" because the "delivery of CBD is a very common experience and really it's just the content within the discipline that's unique" (E6). They explained that commonalties (e.g., questions, data collection tools, etc.) between specialty/subspecialty programs could inform the creation of evaluation matrix frameworks and toolkits by credentialed evaluators to guide program directors, to save valuable time, and to encourage successful evaluations. The "Canadian Evaluation Society e-Institute is [an] online learning platform" (E1) that is available to encourage beginner to in-depth applications of program evaluation. However, with "a common framework . . . you always [have] to get back to 'what are the prime needs? What are the core needs? Who are the stakeholders? What are their needs?'" (E6). An evaluation framework must consider "is it the individual residency program? Is it those who are working at the college level for that discipline? Is it maybe the postgrad executive for that particular faculty in medicine? So, there's probably key clusters and groups [of stakeholders]" (E6). Sometimes, "the common tools are a bit blunt or dull in terms of providing level of specificity and . . . it doesn't quite meet anybody's needs because it may be not quite the question or the information they need. Or, not quite the level specificity" (E6). Nevertheless, "if really there's a recipe to do CBD. You start here. You go to here. Step, step, step. But, really it's just the content is changing by discipline. Perhaps there could be a common approach, methodology, and supporting tools" (E6).

Participants also indicated that despite "having been the person that's written 'how to guides' before. You can develop things and try to give it to people. But what's the old expression, 'you can lead the horse to water but not make him drink?'" (E6). Even with instructions on how to do program evaluation within CBD programs, program evaluators emphasized that resource availability requires personnel to do the evaluation who have evaluation expertise and training. Participants asked "Who's going to do this work? They're so

run off their feet already" (E6), and noted that a staff position is required to "develop their cyclical processes for evaluating their entire program. And then hopefully, they can take it over from there, but I still think they're going to need some resources" (E7). For program evaluation capacity to be initiated, "someone has to take leadership of it . . . and provide some resources or else it can't be done" (E7). Participants explained that "even if they have the time to do it, where they all fall down is they never have the expertise to roll up to analyze the data and report the data" (E6). Therefore, a lack of evaluation "expertise . . . [is] why in many cases you really have to have evaluators doing the evaluation" because evaluators "have the time to do the work and they have the expertise to do the work" (E6). Participants noted that the impact of time and expertise resources is "where the work stops almost every time is on those two points. Do they have people to do it? And then they may almost never have the expertise to see a project and evaluation through all of its phases" (E6). Specifically, "you can give them tools, [program directors] can go out and get the admin support to gather all the stuff," but the investment in evaluation is lost because "[program stakeholders] don't know how to analyze it . . . They don't have the software or the technical expertise" (E6).

Furthermore, program evaluators shared that program directors "face an assessment system or platform that they are unable even to access the data. They have to go to copy and paste the numbers into an excel sheet to be able to see the residents' performance" (E4). This lack of technology infrastructure means that "the data is not used well. It's not represented well and it's not aggregated and visualized in a way that keeps their hands on the pulse of everything that's happening" (E4), which further disrupts analysis. Therefore, it is not realistic "to expect a program to do all of that data collection and analysis [because] they don't have that expertise to do that. So, it works out better when it's centralized and they have centralized support" (E8). Working alongside a credentialed program evaluator can build stakeholders' confidence and empower them because "it's really hard for people who have not done evaluation to do it off the side of their desks" (E1). Participants insisted that "it's a specialized practice and you should work with people who are certified" (E1) to guide program directors by ensuring that they are asking the appropriate questions and gathering data to effectively inform their decision-making processes and not wasting time.

In summary, participants discussed how they can help CBD programs build their capacities to do program evaluation by offering evaluation resources and expertise as guidance in order to save program directors' time.

**Leverage Existing Data.** Participants indicated that they can help CBD programs build their capacities to do program evaluation by maximizing existing data to facilitate the evaluation. For example, program evaluators emphasized "mak[ing] the distinction between assessment and evaluation" because "you have to collect assessment data, but then assessment data is always for an individual trainee's trajectory" (E3). Whereas "a program evaluation . . . is having a programmatic approach to think about the effectiveness of the assessment process [and] the processes within CBD and how they are working together to make the design or the curriculum successful" (E3). 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). Participants also noted that at "the Royal College summits . . . [there is emphasis] around data dashboards and the wealth of the assessment data and how we can use them in program evaluation" (E4). Specifically, "some programs collect around 50,000 assessments per year and . . . there's a huge potential to use this data for improving students, for improving learning, and [for addressing] the problems in general" (E4). For example, programs "collect these huge amounts of assessment data on EPAs . . . the old assessment tools like 145 and ITERs [In-Training Evaluation Reports] and some of the exams" (E4). Participants explained that the dashboard is way to aggregate "all the data sources, including the ITERS, exams, the orals, and . . . they have access to the data all the time" (E4), and they can exploit existing data sources to address program evaluation questions. Moreover, participants suggested making evaluation "more part of whatever [stakeholders] have to do anyways" (E7) because "program evaluation helps . . . get them to stop and think" (E3) about the data that is already collected. However, technology infrastructure needs to be better developed as "clinicians and educators were complaining a lot that they are unable to get to assessment data" (E4), which complicates using existing data for program evaluation.

In summary, participants discussed how they can help CBD programs build their capacities to do program evaluation by leveraging existing data with an adequate technology platform to simplify the program evaluation process.

**Advocate for Clearer Program Evaluation Policy Expectations from the RCPSC.** Participants indicated that they can help CBD programs build their capacities to do program evaluation by building an awareness of program evaluation's power that is supported in policy. Program evaluators specified that "the direction always has to come from the top [RCPSC]. That's a rule" (E9). Participants advocated for explicit and clear expectations from the RCPSC and from PGME departments. Similarly, participants noted that "as an individual evaluator, it's not [our] job to get a sector interested in evaluation—[we] don't need to drum up business . . . [but] I guess evaluators could market to these organizations if they want to" (E1). Program evaluators explained that "every organization has an accountant . . . because they need to demonstrate where their money goes to. But, not all organizations have evaluators . . . [who] show you whether you're spending your money on the right things" (E1). They added that "program managers wouldn't do the accounting work, because they're not trained accountants, but yet they think they can add in evaluation work. It's a specialized area of study and practice" (E1). Participants also suggested that additional communication is required from the Canadian Evaluation Society to inform PGME departments that "there's a whole profession that has education, practice, and professional development to design evaluations that meet stakeholders needs. It takes work." They added that it is problematic that "a lot of people don't even know it exists. So, they don't know what they don't know" (E1). Moreover, participants noted that "competency-based programs [should] also have an evaluator who's competent in their profession" (E1). Developing an understanding of program evaluation is essential because "the postgrad dean . . . prioritized some programs for us to work with. That kind of only worked well if the program director understood program evaluation and knew how to use us" (E3). Therefore, participants insisted that deans and program directors require more information about how to leverage program evaluation experts. For example, some participants explained how "five to eight years later, the postgrad deans, currently, they have a very solid understanding of program evaluation, and they know how they want to interact and use us" (E3). Participants emphasized that the responsibility of the postgrad deans is that "they might see a program that would really benefit from our services, they would put us in touch" (E3). Participants indicated that clearer policy expectations result in better communication between evaluation units and program specialties/subspecialities, and noted that "it's missing . . . that communication piece" because

"even within our department . . . some people were asking, 'where's the program evaluation support unit? Where are you guys? We don't know anything about you'" (E4).

Participants noted that establishing clearer program evaluation policy expectations from the RCPSC would create a common language for communicating, and explained that "program evaluation is a really new concept in postgraduate education. And in fact, they don't even talk about program evaluation. They talk about continuous quality improvement" (E10). Program directors view "continuous quality improvement from the clinical realm, which is a process of 'we have a problem. How do we know it's a problem? . . . And let's implement a solution and do a Plan-Study-Adjust cycle'" (E10). Therefore, program evaluation is "talking a totally new language. We're talking a totally new approach" (E10). As such, "some of [the program directors] do not completely understand what's program evaluation and how it's different from assessment and how it's different from accreditation and quality control" (E4). Participants insisted that it is problematic that the "Royal College has framed it as 'how do you do continuous quality improvement to your program?' So when you talk to a program director, they're not talking about doing program evaluation" (E10). Instead, program directors are "talking about this new standard around continuous quality improvement, which is . . . not a process for figuring out whether you achieve outcomes. It's a process of identifying problems and addressing [them]" (E10). Therefore, "using a systematic approach and talking to the program directors about having a systematic approach to ensuring that they're achieving . . . short-term, medium-term and long-term outcomes? They aren't there yet" (E10). Building a common program evaluation language takes time and "buy-in for program evaluation processes and we're not having those conversations with our program directors right now because it's like talking to a smoker who doesn't want to quit smoking" (E10). However, program evaluators noted that there is never a "right" time to quit smoking, and program evaluation conversations will occur when the RCPSC, PGME departments, and program directors understand its value as documented in policy.

In summary, participants discussed how they can help CBD programs build their capacities to do program evaluation by advocating for clearer program evaluation policy expectations from the RCPSC.

### *Research Question 6b: How Can Program Evaluators Help CBD Programs Build Their Capacities to Use Evaluation Findings?*

In the sections that follow, I describe how the participants can help CBD programs build their capacities to use evaluation findings. Specifically, program evaluators can: (a) recommend that program evaluation is 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 Evaluation Is Included at the Beginning of Program Design and Implementation.** Participants indicated that they can help CBD programs build their capacities to use evaluation findings by recommending that program evaluation is incorporated at the beginning of program design and implementation. Program evaluators noted that "the best model of evaluation is when you have the evaluator at the planning table" (E6). When invited to PGME or program committee meetings, evaluators can "hear the conversations, hear the context, hear the direction, the shape of the program, where it's going, timetable" (E6). Moreover, when evaluators are "involved up front rather than after the fact . . . they can kind of start formulating their thinking around what is the best strategy" (E6). However, participants noted that one challenge in evaluating CBD is that program stakeholders are "absolutely mired and overwhelmed . . . in just trying to do [CBD], that they are not thinking about evaluation yet" (E6). For example, program directors are asking "How do we do [CBD] in our program? We got to get the faculty trained out, but we got to get this this software platform organized, so we can do our assessments" (E6). With the workload that accompanies CBD implementation, a program director's "mind is not . . . try[ing] to understand how we're doing with all this [because] . . . we're really in the middle of rollout" (E6).

Nevertheless, participants insisted that implementation rollout could be smoother with appropriate evaluative planning. They noted that conducting program evaluations as an afterthought "is really unfortunate because . . . what we have in terms of findings, some of what we're seeing could be related to program design" (E2). Participants added that "if an extensive

program evaluation is done after the fact, I'm just not sure how much use of evaluation findings would actually occur" (E2). Furthermore, participants "doubt[ed] that a lot of use of the findings would occur because, again, of how much money and how much resources was put into the rollout of the program" (E2). In other words, participants thought it was unlikely that more financial resources would be applied to undo recent efforts.

In summary, participants discussed how they can help CBD programs build their capacities to use evaluation findings by recommending that program evaluation is included at the beginning of program design and implementation—even if its inclusion is only for evaluators listen to the conversations and to avoid overloading program directors.

**Include the Stakeholders Who Will Be Using the Findings in the Evaluation Itself.** Participants indicated that they can help CBD programs build their capacities to use evaluation findings by including the stakeholders who will later be using the findings during the evaluation process. Program evaluators insisted that stakeholders "need to have some level of involvement in the evaluation . . . we could have different levels of involvement, even if we're just talking about some shared decision making, versus a very high level of engagement in evaluation from beginning to end" (E2). Participants emphasized the importance of clarifying the degree of each stakeholder's participation and their role in the evaluation. Specifically, they identified that "there just needs to be some level of involvement because then it allows those individuals to understand what is occurring. Why is it occurring?" (E2). Ensuring that stakeholders understand "the 'why' behind it. It is always such a big thing for buy-in and for people to understand why something is important" (E2). When stakeholders understand "Why are we asking these evaluation questions? Why are we collecting this data? What are we going to use it for? That understanding, truly becomes developed and ingrained and maintained through some level [of] involvement" (E2) and it increases the likeliness of the findings being used for program improvement. Participants noted that "even if it's just shared decision making or even if it's just always keeping stakeholders up to date" (E2), engagement in the process demonstrates commitment to improving an area of the program and can increase the likeliness of the findings being used.

However, some participants commented that "knowing that not all programs are at that space yet, we have opted to wait before putting a programmatic approach into asking about CBD specifically" (E3). They also noted that "the curriculum is the curriculum, the clinical services

are the clinical services and the clinical education like it's [the] same" (E3). The transition in PGME programs is that "CBD is a lot of the assessment part of the competency like how they get at the competency. The CBD uses the design of which you deliver the curriculum" (E3). Therefore, participants explained that "what the residents actually do day-to-day is not changing that much. So, when we're asking them questions about their training experience in relation to the curriculum, first of all, a lot of them don't even know" (E3). Answering questions, such as, "do you know that your program is doing CBD and how has that changed?" is challenging for residents to answer because "they can only respond based on what they interact with their everyday" (E3). As such, program evaluators emphasized the need to include appropriate stakeholders in an evaluation.

In summary, participants discussed how they can help CBD programs build their capacities to use evaluation findings by including the stakeholders who will be using the findings in the evaluation process.

**Focus Stakeholder Questions on Meaningful Program Areas Over Which They Hold Authority to Make Changes.** Participants indicated that they can help CBD programs build their capacities to use evaluation findings by focusing stakeholder questions on areas that stakeholders care about and over which they hold power to alter. Program evaluators explained that "what made it easier is that people were genuinely interested in finding out these answers themselves" (E10). Participants shared that asking questions that stakeholders care about 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). Specifically, "you need to have that buy-in from the get-go—not midway through, not at the end. Those are not the time points to be trying to establish buy-in. It has to be before [the evaluation] even starts" (E2). To facilitate this, participants suggested to "lay out your [authority appropriate] evaluation questions and [then] you develop a design that will help you answer those questions" (E1).

Furthermore, program evaluators indicated that questions should investigate areas over which stakeholders hold authority to make changes, and they will be different at various hierarchical levels (i.e., the RCPSC, PGME, and program director). Specifically, at the RCPSC level, participants suggested that purposeful evaluation questions stakeholders can document include "is this really achieving the end result that we said it was? Do we need to steer the ship

[differently]? Or, are we even on the right ship?" (E6). They added that "the investment for evaluation would be very worthy. And there's probably enough combination of what they call the big and the small programs to do some fairly robust evaluation" (E6). Participants also emphasized working with the PGME department above "the program director level because most of the evaluations of that [we] do is more at the chair of the department or physician level where they want something to happen" (E5). Participants noted that program evaluators can listen to "what are the questions that these leaders at this table are formulating right now that they want to have addressed?" and insisted "that kind of model is always best" (E6).

At the program director level, program evaluators can "start with some foundational questions . . . on the implementation delivery domain, and in the outcome domain" (E6). Focusing on questions about implementation and outcomes "will drive improvements . . . year by year to keep moving forward with CBD" (E6). Participants recommended to "start with some very foundational core evaluation questions . . . that doesn't have to be too detailed" to support this "big organizational transformation" (E6). They also noted that "one of the main findings was the concurrent implementation of a bunch of other things that were happening at the same time that were out of the control of the program director . . . so, while [the program director] would like to make those changes . . . those changes are actually out of their control" (E10). Moreover, participants explained that "one of the biggest challenges when you think about [program evaluation] being implemented at the program director level is many of the challenges were system problems" (E10). For example, the problem was "the new learning management system, the new accreditation standards, and CBD all at the same time. So, there was a lot of system level challenges that impeded the implementation" (E10). In this context, "the learnings are more for the postgraduate office. The learnings are more for the Royal College" (E10). However, participants noted that they "can't tell you how the Royal College may or may not choose to use the findings" (E10).

In summary, program evaluators discussed how they can help CBD programs build their capacities to use evaluation findings by focusing 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.** Participants indicated that they can help CBD programs build their capacities to use evaluation findings by using program evaluation approaches that are

appropriate for the CBD implementation context. For example, program evaluators explained that using "a realist evaluation recognizes that problems with implementation are not the responsibility of an individual" (E10). Participants noted that applying a realist approach removed blame associated with implementation failures and facilitated inquiry towards improvement. Participants also noted that when evaluation becomes more about learning and improvement and less about criticizing, program directors are more likely to use the findings.

In addition, participants stated that other program evaluation approaches that encourage the use of findings are participatory and collaborative evaluation approaches. Specifically, participants explained that they are "a huge proponent for participatory and collaborative evaluation approaches [when] talking about the field of medical education" (E2). They commented that "when we ask the questions that they want to ask, when we listen to what are their needs, and try to find answers for them . . . it goes back to the idea of participatory evaluation" (E4). Similarly, participants shared that "it really comes down to a collaborative or participatory model for evaluation. I think that's really where leadership really feels that they are part of and invested in the process that usually yields the best uptake" (E6). They indicated that "the high stakes of these training programs and the widespread impact that we're talking about . . . these are not programs that only impact the learners. These are programs that impact patients, families, communities" (E2), which makes it even more important that the findings are used for improvement. Thus, a "practical participatory [approach] is something that marries much more naturally in terms of an approach within medical education" (E2). Specifically, "when we are conducting research, we need to be very practical with our data collection methods and . . . do what works just because the context of health and medicine is so unique, is so complex" (E2). Therefore, to make use of program evaluation findings, program evaluators "need to be flexible and do what works within the context. Which leads me back to why practical works so well and . . . being flexible with regards to what that involvement looks like and what the level of involvement" (E2). Participants emphasized that they "always recommend . . . the practical participatory way" (E2).

Finally, to encourage the use of findings, some participants noted that they applied the New World Kirkpatrick Model, explaining that program evaluators "always talk about only logic models and the [New World] Kirkpatrick Model. They are the two most extensively used models within medical education" (E4). These participants described how the New World Kirkpatrick

Model explores the reactions, learning, behaviour, and results. However, participants asserted that "we need to make sure that we are really doing CBD in the first place" (E4) before using the findings to measure the impact of training. An appropriate CBD evaluation approach should be evaluating implementation fidelity "because if the theory is good, but the implementation is not working [according to] the theory . . . the theory should not be accountable for the outcome that we have" (E4).

In summary, participants discussed how they can help CBD programs build their capacities to use evaluation findings by encouraging the use of program evaluation approaches that are appropriate to the CBD implementation context.

**Interpret the Findings Alongside the Stakeholders and Co-Develop a Manageable Follow-Up Plan.** Participants indicated that they can help CBD programs build their capacities to use evaluation findings by interpreting the findings with the stakeholders and co-creating a manageable follow-up plan. To ensure that stakeholders understand findings, participants insisted that program evaluators should participate in the summary and interpretation of quantitative and qualitative data collection and analysis. For example, participants explained that they "can facilitate what people call 'data parties' or 'meaning making sessions' or 'action planning sessions' using common framework like the so-called adaptive action framework" (E1). Some participants facilitated a findings discussion to consider three questions: "what's the data? So, what is it telling us? And now what?" (E1). These program evaluators noted that when the findings are clearly understood and dissected alongside a program evaluator, stakeholders who were passionately involved are more likely to use the findings. Participants also suggested to "have some of the high users so, the people who do this well . . . train others—like train the trainer . . . [because] if they see success, it breeds success" (E5). Participants indicated that stakeholders who use findings well can share their successes with other programs, which can create an evaluation culture that uses the findings.

Participants also emphasized the importance of co-developing a follow-up plan that is manageable for stakeholders. They specified that a program evaluator can "get them to think about risk management response plans to evaluations where, again, they look at the evaluation, and the findings, and they decide what they're going to do about it" (E1). Program evaluators can break down a plan with stakeholders to decide "how to move forward in actioning the findings" (E1). The idea of a follow-up plan or cycle makes evaluation "integrated as a normal, natural

practice and not something that's an add on" (E2). Participants added that it was helpful to "formalize the follow up, the tracking of . . . in what ways or to what extent have [the recommendations] been implemented and what are the challenges?" (E3). This way of institutionalizing an evaluation cycle establishes that in "the next round of evaluation we can kind of follow-up on that . . . and ensuring that they are being acted upon" (E3). They explained that "we give [the recommendations] back to the program committees . . . [and] we're going to follow-up. And so just having that reminder" (E3). Participants insisted that by co-creating a plan of action with stakeholders, "they have it incorporated, established in their committee agenda, a time where they actually come back and look at these recommendations" (E3). Specifically, "it's something that is feasible for them to do . . . [based on] the stepwise ways in which they can address it" (E3).

In summary, participants discussed how they can help CBD programs build their capacities to use evaluation findings by interpreting the findings alongside the stakeholders and co-developing a manageable follow-up plan.

**Encourage Programs to Share Findings, Which Establishes an Accountability Cycle to Make Changes Based on Findings.** Participants indicated that they can help CBD programs build their capacities to use evaluation findings by pushing programs to share findings, which develops an accountability cycle in order to make adjustments based on findings. Program evaluators shared that although an iterative cycle of evaluative review "isn't always robust or really an authentic, embedded function within the organization" (E1), sharing findings is one way to keep stakeholders accountable for applying the findings because there may be a future check-in (E1). Specifically, participants commented that the dispersion of findings should include "regular touch bases, and whatever format that knowledge mobilization looks like to work most efficiently with transmitting these messages to faculty members and whomever else" (E2). Participants indicated that one person can lead this process, but it involves "making sure that the faculty are aware of what's happening, feeding them back results" and it helps stakeholders to "actually see . . . value or they're at least aware of it" (E7). Some program evaluation findings have "already been presented at the program evaluation summits that [the] Royal College had, and the findings have been shared in a written report with postgraduate programs" (E10). However, participants acknowledged that the findings were "maybe not

[shared with] the residents, but certainly the program directors and program administrators" (E10).

Program evaluators also noted that "we want to get themes. We want to get a general sense of what's working, so that other programs that haven't launched yet [can] learn from it" (E10). Participants explained that sharing areas of implementation successes and lessons learned from rigorous and documented program evaluation is a way of paying it forward to program specialties and subspecialties, building a stronger PGME community, and establishing a more effective Canadian PGME system. Program evaluators noted that "with regards to say patients and families and community organizations as they continue to be involved then, they develop the mindset . . . to hold our programs accountable . . . which falls directly in line with social accountability" (E2). Similarly, sharing findings established "ongoing annual [and] biannual program evaluation services . . . [and] reporting . . . [which] speaks to accountability and social accountability mandates" (E6). According to participants, the reports encourage programs to "be accountable to different types of organizations . . . federal, provincial, municipal, governments, NGOs, all those kinds of things in the field of education, a lot in health care" (E6). Participants insisted that sharing findings made stakeholders accountable for how the findings are making a difference, which would be reviewed in the future (i.e., a check-in cycle to see how the findings are being used). Furthermore, participants emphasized "that curricular reform is an iterative process. So, we always understood that it would be an ongoing effort to truly transition" (E8).

However, depending on the research approach (i.e., a realist evaluation), the findings may not be integrated into an intentional programmatic cycle because "it's not an ongoing evaluation, like you would see if you used a different kind of approach to evaluation" (E10). Some participants explained that they "evaluated the implementation of CBD . . . as a scholarly project . . . [and] didn't plan an ongoing evaluation of CBD" (E10). The purpose of engaging in program evaluation this "way [was] so that we'd have lessons learned for next time as we did, because we want to know what works under what circumstances for who and why" (E10). In these cases, the findings were used to offer "a lesson learned next time you go to do an evaluation" and contributed to knowledge about how to conduct program evaluation in medical education (E10). Program evaluators noted that sharing findings about research on program evaluation also established accountability among program evaluators to use these findings the next time they support a specialty/subspecialty program evaluation.

In summary, participants discussed how they can help CBD programs build their capacities to use evaluation findings by encouraging programs to share findings, which establishes an accountability cycle in order to make changes based on findings.

**Summary for Phase 3**

Overall, 12 program evaluators consented to be interviewed for Phase 3 of my study about how they are currently or could be supporting CBD programs in program evaluation and how they can help CBD programs build their capacities to do program evaluation and use evaluation findings.

Participants discussed how they are currently supporting or could support CBD programs in program evaluation. Specifically, they are or could be: (a) responding in a reactive way as a temporary and external evaluation consultant; (b) collaborating with interested stakeholders as internal evaluation experts from a 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.

Participants also discussed how they can help CBD programs build their capacities to do program evaluation. Specifically, they 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.

Finally, participants discussed how they can help CBD programs build their capacities to use evaluation findings. Specifically, they can: (a) recommend that program evaluation is 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.

# Chapter 7: Discussion

In this chapter, I position the findings of my thesis within the existing literature and examine my study's contributions. In the first section, I summarize the main findings from each thesis section. In the second section, I integrate my findings within the research literature on program evaluation in medical education, common program evaluation approaches in medical education, and building evaluation capacity to do program evaluation and to use program evaluation findings in medical education. Finally, in the third section, I clarify how this study contributes to how to do quality assurance on CBD and offers practical and methodological contributions to scholars conducting research on program evaluation in medical education and to evaluators evaluating programs in medical education.

## Summary of the Main Findings

Through this research, I discovered some interesting trends in:

- the extent that and the ways in which CBD programs are engaging in program evaluation;
- the reasons for CBD programs engaging or not engaging in program evaluation;
- the actual and potential positive and negative consequences of these programs engaging in program evaluation;
- the strategies to support CBD programs in building their capacities to do evaluation and use evaluation findings;
- the ways in which program evaluators are supporting CBD programs in program evaluation; and
- the strategies to support program evaluators helping CBD programs in building their capacities to do evaluation and use findings.

Specifically, from the Phase 1 survey of program directors, I found that most of the respondents identified that their CBD program engages in program evaluation. Of the respondents whose CBD programs engage in program evaluation, more than half of the respondents indicated that they have either "never" or "rarely" worked with a program evaluator. Most of the respondents indicated that they invite stakeholders to participate in program evaluations either "frequently" or "always". More than half of the respondents answered that they "rarely" or "sometimes" interpret quantitative program evaluation data. Furthermore, when reflecting about the purpose of their CBD program evaluation more than three quarters of the

respondents indicated that they “sometimes” or “frequently” evaluate whether their program has been implemented as intended. Most respondents indicated that their program does not have an employee whose primary responsibility is program evaluation. Most respondents indicated that their program does not receive funding for program evaluation.

The few respondents who indicated that their CBD programs do not engage in program evaluation clarified some of the reasons for not engaging in program evaluation. Over half of respondents indicated that their program does not engage in program evaluation because the program has no personnel and no funding to do program evaluation. A majority of respondents indicated that their program does not engage in program evaluation because the program does not know how to do program evaluation.

Respondents also indicated that engagement in program evaluation is limited due to cost, time, and expertise constraints. The collection, analysis, and evaluation of program data are low priorities because program directors lack time. Moreover, when program evaluation findings are acquired, there is a lack of power to make changes because the PGME office and the RCPSC make decisions. Thus, some program director stakeholders do not feel empowered to engage in program evaluation. Furthermore, some of the suggested program revisions are not possible because of budget constraints. Some program directors indicated that they are unclear about what program evaluation means. Finally, certain program directors indicated that they had started to assess their programs in an attempt to encourage ongoing quality improvement, but feel that a standardized approach is required.

From the Phase 2 interviews with program directors, I discovered that CBD programs are struggling to engage in program evaluation because of limited available resources and buy-in. The program director participants who engage in program evaluation indicated that they are using ad hoc evaluation methods and team-based formats. Participants suggested strategies to build their capacities to do program evaluation, including developing expertise in program evaluation abilities by acquiring program evaluation resources based on clear program evaluation expectations from the RCPSC, the PGME office, and within their specialties. Participants also indicated that they can build their capacities to use evaluation findings by asking meaningful questions that are appropriate to the stage of the CBD program, which can help to establish program evaluation buy-in and increase the likeliness that the findings will be used. Finally, program directors suggested that they can share findings with faculty, residents, and other

programs in ways that encourage use and view evaluation as helpful for the accreditation process, so program directors are more likely to use the findings.

In the Phase 3 interviews with program evaluators, I found that program evaluators are supporting CBD programs in program evaluation by responding in a reactive way as temporary, external evaluation consultants, collaborating with interested stakeholders from an internal position within an evaluation unit at a university, and leading evaluation approaches that are manageable for program directors with limited time. While acting as either an external consultant, an internal collaborator, or a lead evaluator, the program evaluator participants indicated that they are using methods that complement the busy PGME environment without requiring additional resources, using evaluation processes that lead to the publication of findings, and offering professional development sessions about program evaluation to support accreditation. Participants also suggested that they can help CBD programs build their capacities to do program evaluation by applying a participatory approach, offering evaluation resources and expertise, leveraging existing data, and advocating for clearer program evaluation policy expectations from the RCPSC. Finally, participants suggested that they can help CBD programs build their capacities to use evaluation findings by recommending that program evaluation is included at the beginning of the process, including the stakeholders who will be using the findings, guiding stakeholder questions, encouraging the use of appropriate program evaluation approaches, interpreting the findings alongside the stakeholders, and encouraging programs to share their findings.

**Integration of the Findings with the Literature**

In this study, I documented the extent to and the ways in which CBD programs are engaging in program evaluation, the reasons why CBD programs are engaging or not engaging in program evaluation, and the strategies program directors and program evaluators are using to support CBD programs in building their capacities to do evaluation and use findings. This documentation provides empirical research on program evaluation in medical education and information for evaluators who are embarking on evaluations of medical education programs. To make sense of my findings, I integrate them within the larger bodies of literature on program evaluation in medical education and building capacity to do and use evaluation.

In the literature review in Chapter 2, I presented an overview of the purposes of program evaluation in medical education, common program evaluation approaches in medical education,

support from program evaluators to evaluate CBD programs, and ECB in medical education. I argued that specialty/subspecialty program evaluations are required to learn more about the recent PGME transition to CBME, or CBD, as it is described in Canadian medical education. Conducting CBD program evaluation is necessary because, despite the transition to CBD, there is minimal evidence to support whether CBD develops increasingly prepared physicians or improves patient outcomes compared to traditional medical education (Van Melle, Gruppen, et al., 2017). It is also important to evaluate CBD programs because the transition to CBD required a resource-intensive investment (i.e., time, energy, financial, and human resources) by the RCPSC, PGME departments, program directors, and faculty—and it is essential to ensure that residents continue to receive high-quality medical education.

To understand how CBD is operationalized within specific residency programs and how it contributes to patient, faculty, and learner outcomes, there is a need to engage in program evaluation. Patton (1997) characterized program evaluation as systematically collecting information about a program's activities, characteristics, and outcomes to make informed decisions about the program and to improve program effectiveness. Through this study, I found that despite program directors' interest in engaging in program evaluation, certain respondents are not engaging in program evaluation because of limited time and a lack of understanding about how to get started with evaluating their programs.

Cook (2010) claimed that program evaluation is essential because it provides vital information about: (a) the types of educational programs needed, (b) whether programs are being implemented as intended, (c) how programs can be improved, and (d) whether certain programs should continue. In this study, the program directors noted their interest in using program evaluation to learn more about the transition to CBD, and particularly the outcomes. However, the program evaluators noted that before evaluating the outcomes, it is essential to learn more about whether programs are being implemented as intended, because an error in implementation could result in problematic outcomes. In other words, a program theory (i.e., CBD) cannot be held responsible for unachieved outcomes if the program was not implemented appropriately. The program evaluators stated that evaluation of implementation is required in order to know more about whether programs are appropriately implementing CBD. Van Melle et al. (2019) created the Core Components Framework, which offers a common structure to guide evaluation on whether CBD is being implemented as intended. In this study, one program evaluator used

this framework to structure their evaluation as a methodology leading to publication of findings. Van Melle et al. (2019) claimed that the intentional linking of theory and practice in the CBD Core Components Framework provides a significant step towards a deeper understanding of the conditions in which CBD most effectively enhances patient care outcomes. However, this study highlights a general lack of intentional linking of theory and practice in program evaluation of CBD programs.

Vassar et al. (2010) found that program evaluation in medical education often lacks proper forethought, structure, and methodology. My findings show that program directors are using ad hoc methods to engage in program evaluation because of limited time, limited human resources, and no additional funding. Mertens and Wilson (2012) argued that an investigation about the extent to which program evaluation is occurring can help to improve program evaluation in CBD and clarify the relationships (or lack thereof) between educational objectives and outcomes, explore the use of evaluation findings to inform program decision making, and determine programs' values or worth. In my study, program directors emphasized informal discussions with other faculty members as the basis for their decisions. However, they were interested in how program evaluation could offer concrete evidence and establish a use of findings cycle to follow up on whether their decision making resulted in the intended changes. Rossi et al. (2019) claimed that to guide program evaluation, the participating stakeholders (e.g., program director, faculty, program administration, residents, clinical site-supervisors, etc.) can develop major evaluation questions centering on, for example, the need for the program, program theory and design, processes, outcomes, efficiency, or impact. However, in this study, program evaluators advocated for prompting evaluation questions that are appropriate to the hierarchical level. In other words, the RCPSC, PGME departments, and program directors will have different program evaluation questions because they will use the findings to apply changes that are within their control.

### *Purposes of Program Evaluation in Medical Education*

The literature explained that stakeholders are interested in doing program evaluation in medical education for a variety of purposes. For example, Cousins et al. (2015) noted that stakeholders may decide to conduct program evaluations for accountability or learning purposes. In my study, program directors indicated that they wanted to engage in program evaluation mainly for accountability purposes, either in preparation for or following accreditation from the

RCPSC. However, some program directors demonstrated self-motivation to learn more about program evaluation to improve their programs. Specifically, the sample of program director participants who agreed to be interviewed for Phase 2 are likely those who are most interested in program evaluation. Notably, the program directors from Phase 2 indicated that conducting program evaluation is not realistic for them without additional resources (e.g., time, financial, human, and technological infrastructure). Thus, I am skeptical about the extent of program evaluation that is actually occurring in Canadian specialty/subspecialty programs, despite the Canadian Residency Accreditation Consortium (CanRAC) Accreditation Element 9.1.1, which states: "There is a systematic process to regularly review and improve the residency program" (CanRAC, 2020, p. 18).

Clarke (2006) explained that, in terms of accountability, stakeholders can evaluate program effectiveness, identify best practices, and validate their programs within financially competitive climates. In this study, program directors noted the opportunity to evaluate program effectiveness by reviewing whether the larger amount of ongoing feedback helped to identify and support struggling residents sooner. Alkin and Christie (2012) noted that there are three types of accountability in program evaluation, including (a) goal accountability (i.e., program goals have been developed; is of particular interest to governing boards and upper-level management), (b) process accountability (i.e., examines the implementation and appropriateness of procedures used to obtain program goals; is the responsibility of faculty and learners), and (c) outcome accountability (i.e., the degree to which goals have been achieved; is also the responsibility of faculty and learners). In the literature, medical educators and learners have documented short-, medium-, and long-term outcomes attributable to programs (Van Melle, Frank, et al., 2017). However, in my study, program evaluators insisted that process accountability must be reviewed prior to attempting to reach outcome accountability. That is, the implementation and appropriateness of the procedures used to obtain program goals (as determined by the RCPSC) must be assessed before evaluating or making decisions about the degree to which goals have been achieved. The literature also suggested that program evaluation can provide stakeholders with insight into the "black box" of program functioning (Funnell & Rogers, 2011). My findings indicate that program directors are interested in learning about the effectiveness of CBD transition, but program evaluators recommend clarification on whether learners' outcomes are attributable to the program theory, objectives, and/or implementation. In 2019, the RCPSC

completed two program evaluation studies on the implementation of the 2017 and 2018 CBD launch disciplines and the level of readiness of the 2019 CBD launch disciplines (RCPSC, 2019). My findings agree with the RCPSC's findings that challenges include the pace of culture change affecting aspects of CBD (e.g., giving and receiving feedback), workload increase, the impact of CBD on stress and well-being, training and information, and electronic platform creation (RCPSC, 2019). Although the goal of the RCPSC evaluations was that medical education disciplines commencing their CBD journeys would be able to learn from the findings and use them in the planning, implementation, and evaluation of their own programs, one program director in my study noted that the RCPSC evaluations leveraged financial and human resources to do program evaluation that are not available to other program directors. Thus, the RCPSC is asking program directors to do program evaluation, but without providing additional support or resources.

Regarding the use of program evaluation for learning purposes within medical education, Torres et al. (2005) claimed that by actively participating in program evaluation, individuals can increase their competencies in data collection, analysis, and communication (behaviourism). However, in this study, program evaluators noted that where program evaluation often fails is in the data analysis because program directors often lack data analysis expertise or do not have time to analyze the data. The program evaluators who discussed successful program evaluations of CBD specialties emphasized the importance of involving program directors at the beginning of the evaluation process but involving them less in the data collection and data analysis because of the need for the expertise of a data analysis team. Torres et al. (2005) also explained that through participation in program evaluation, individuals can begin to understand issues more completely and gain greater appreciation of issues (constructivism) or increase their abilities to think differently about program problems or issues by applying new cognitive clues (cognitive learning). In this study, program directors noted that they are applying a team-based format to collaboratively, yet informally, explore program issues or problems to obtain a complex understanding of their program.

### *Common Program Evaluation Approaches in Medical Education*

Regarding the common program evaluation approaches in medical education, the literature revealed a need for further systematic program evaluations of CBD programs and provided examples of the types of evaluation approaches that programs may use. As medical

educators consider how to evaluate their CBD programs, the literature emphasized that it is essential to recognize stakeholders' evaluation needs and to adopt appropriate evaluation approaches (Cook, 2010). Program evaluation researchers within medical education have suggested that objectives-oriented, process-oriented, and participant-oriented approaches work well for evaluating medical education programs (Cook, 2010).

In objectives-oriented approaches, instructional goals are defined at the beginning of the activity and evaluated at the end to determine if the specified goals have been met and if there is a cause–effect linear relationship in the program's elements (Stufflebeam & Shinkfield, 2007). In the literature, contribution analysis exemplifies an objectives-oriented approach that is theory-based for evaluating the performance of programs, and is ideal for the evaluation of medical education programs. The New World Kirkpatrick Model is another objectives-oriented evaluation approach that includes four levels (i.e., reaction, learning, behaviour, and results; Kirkpatrick Partners, 2020). In this study, program evaluators noted that the New World Kirkpatrick Model is an approach that is often used in medical education, but they also acknowledged that the transition to CBD lends itself more to a process-oriented approach to determine the fidelity of implementation. An objectives-oriented approach may be more useful after appropriate implementation has been established to determine whether CBD better prepares residents for practice and improves health care for patients.

Process-oriented approaches apply general systems theory to determine if a programming need exists, to outline the most appropriate way to satisfy that need, and to track the development process during implementation (Cook, 2010). In the literature, the CIPP model is a process-oriented approach to program evaluation that focuses on program improvement instead of attempting to prove an aspect about the program (Stufflebeam & Shinkfield, 2007). However, in this study, program evaluators never mentioned the CIPP model. According to CanRAC's General Standards of Accreditation for Residency Programs, the Continuous Improvement domain includes standards to establish a culture that mimics the continuous improvement cycle (i.e., Plan–Do–Study–Act) and Element 9.1 expects that "the residency program committee systematically reviews and improves the quality of the residency program" (CanRAC, 2020, p. 18). However, in this study, program evaluators noted the problem with language discrepancies and associating program evaluation with words such as "continuous improvement," "quality improvement," or "an evaluation of" (CanRAC, 2020). Instead, the

program evaluators advocated for being explicit and using the words “program evaluation.” Furthermore, program evaluators indicated that Plan–Do–Study–Act is how physicians understand improvement by identifying a problem and making a plan for remediation, but it is not a process to systematically review a program and use of the term can create confusion among program directors who are interested in engaging in professional development about program evaluation.

In their work, Hall et al. (2020) showed that rapid evaluations provide insights into the successes and challenges of CBD implementation. In this study, some program evaluators used a continual improvement process, or developmental evaluation process, because the flexibility allowed for stakeholders to make changes in real time and incorporate them into what was already being done. Specifically, the program evaluators highlighted their use of a rapid cycle evaluation approach to review the implementation of interventions and to lead an evaluation that is manageable for program directors with limited time. One internal evaluation unit led a rapid cycle evaluation when specialty/subspecialty programs transitioned to CBD, and it informed their adjustments. This approach was possible because the CBD faculty lead valued program evaluation and advocated for allocated funding to support data collection, analysis, and report draft write-up, per the Core Components Framework. These program evaluators managed the process so it was not labour intensive for program directors, which made all the difference, and they standardized the methodology, allowing for the publication of findings. The program evaluators explained that while the program directors were genuinely interested in learning more about their programs and how to improve them, they were also enticed by the publication of findings.

Another process-oriented approach that program evaluators discussed during Phase 3 of this study is a realist evaluation of the implementation of CBD. According to the program evaluators, program directors appreciated that the objective of this approach was to explore “What works? For whom? In what circumstances? And, why?” A realist evaluation is an exploratory approach, where evaluators look at the interrelationship between the context in which the implementation is happening and the resulting mechanisms and outcomes. Within medical education, there is a need for practical and detailed understandings about which educational interventions work, for whom, in what circumstances, and why (Van Melle, Frank, et al., 2017). In this study, the realist evaluation that program evaluators discussed did not have a

utilization focused evaluation design that measured something over time, made changes, and waited to see if it resulted in improvement. Instead, implementation was contextually bound to learn more about what is happening, when things are working, and why they are working, using complexity theory that acknowledges that just because something works in anesthesia does not mean it will work in internal medicine.

Participant-oriented approaches address how stakeholders involved in a program view it, use an ongoing cycle of data collection and interpretation, and apply triangulation from multiple perspectives (Cook, 2010). PE is a participant-oriented approach that is commonly advocated for in medical education because it enriches a sense of ownership over the evaluation, embeds evaluative thinking into an organization's culture, and enhances the use of evaluation findings (Patton, 2012). In this study, program evaluators noted that including program directors in the evaluation supports buy-in because they prompt the program directors with questions that are meaningful. Furthermore, based on its collaborative emphasis between trained program evaluators and stakeholders, PE is useful within CBD because it can provide medical education decision-makers with access to real-time evaluation data and build necessary evaluation capacity among individuals who must maintain continuing evaluation activities (Moreau, 2017). In this study, program directors advocated for additional resources to develop their expertise about how to engage in program evaluation.

In the literature, PE includes two streams: practical participatory evaluation (P-PE) and transformative participatory evaluation (T-PE). The primary objective of P-PE is to facilitate program decision making and problem solving, and to apply evaluation findings and processes (Cousins, 2005). It is geared towards formative program evaluations, which improve stakeholders' understanding and implementation of their programs (Cousins & Earl, 1992). Typically, P-PE only involves primary stakeholders who are in leadership roles within programs and can thus use the evaluation findings to facilitate programming changes. In this study, program evaluators advocated for the use of P-PE because of the busy environment of the program director and faculty, as many of these stakeholders are not only teaching faculty but also practicing physicians. The primary objective of T-PE is to empower community groups that are oppressed by dominating groups, and because it is stakeholder-driven, T-PE allows the evaluator to take on a purely facilitative role (Cousins & Whitmore, 1998). However, according to my findings, T-PE is not practical for program evaluation in medical education, because program

directors' limited time and lack of expertise in program evaluation requires that a program evaluator take the lead, at least until further capacity is developed. This finding aligns with Gruppen et al. (2018), who found that it is unrealistic to expect CBD residents and faculty to recognize and apply constructive program evaluation feedback without support or skill development. My findings also agree with Oandasan et al. (2020), who suggested that a professional evaluator (i.e., through the Canadian Evaluation Society; AEA; International Organization for Cooperation in Evaluation) be enlisted to support the planning and implementation of a CBD program evaluation.

***Capacities to Do Program Evaluation***

In contrast to an imposed, one-time evaluation, ECB applies evaluation theory, practice, and policy to facilitate an ongoing and sustainable learning community (Preskill, 2008). In the literature, minimal attention has been given to building program evaluation capacity within medical education programs, such as augmented motivation, knowledge and skills to improve programs, and organizations' abilities to do evaluation and to use evaluation findings (Labin et al., 2012). The literature has described two types of ECB: (a) direct ECB that involves planned ECB activities within and outside of evaluation projects, such as training on data analysis; and (b) indirect ECB that is the product of stakeholder involvement in evaluation processes that lead to building evaluation knowledge (Cousins et al., 2004). In this study, program evaluators stated that they offer professional development about program evaluation to increase program director, faculty, and resident knowledge (i.e., direct ECB). However, my findings also suggest that indirect ECB, as the learning function of stakeholder participation that builds their understanding of program evaluation and processes, is a way of learning by doing—and, I suggest, a more powerful way of demonstrating the value of program evaluation. Instead of focusing solely on program evaluation findings, it changes individuals' technical abilities (i.e., to make sense of findings) and their perspectives of evaluation.

The capacity to do program evaluation refers to the knowledge and skills needed to carry out evaluation activities. Cousins et al. (2008) suggested that these evaluation abilities include evaluation planning, standards of professional evaluation practice, instrument development, data collection, analysis, reporting, and follow-up. However, I found that program directors do not know whether they are asking appropriate questions in their data collection instruments, do not have the technological infrastructure to adequately analyze the existing data, and do not follow

up on findings or changes that they have made. These program directors' challenges to do program evaluation echo Bourgeois and Cousins (2013), who emphasized the importance of human resources, organizational resources, and evaluation planning and activities.

In my study, program directors indicated that the lack of human resources for program evaluation was impacted by minimal staffing of evaluation positions within their PGME organization, lack of knowledge on whether there was an evaluation expert in their PGME, and lack of clarity on whether there were professional development activities. However, one program director indicated that there was an evaluation expert to support CBD implementation and questioned whether he was the only one accessing this human resource because the feedback and support that he was receiving was outstanding (i.e., either the evaluation expert was amazing to be able to support over 60 specialty/subspecialty programs or not many programs were leveraging her expertise).

Organizational resources for program evaluation are based on the evaluation budget's stability, ongoing data collection for performance measurement systems, and organizational infrastructure that supports evaluation policies and governance structure (Bourgeois & Cousins, 2013). In my study, program evaluators who held internal evaluation unit positions shared that their positions were not indefinite. This instability in universities' evaluation budgets reflects a lack of evaluation prioritization. Although the program directors explained that ongoing data collection for performance measurement systems are in place, the collected data is challenging to access and program directors often resort to copying and pasting the data into excel spreadsheets. Moreover, program directors indicated that further clarification that specifies evaluation policies is required in the organizational infrastructure. Program directors required additional information about what, the extent, and how often evaluation should be occurring within their programs. The RCPSC level, PGME department level, and program director level all require clear evaluation policies if program evaluation is going to be applied.

Evaluation planning and activities involve external supports from professional associations and the distribution of evaluation results to members of the unit (Bourgeois & Cousins, 2013). In my study, program directors noted that they do not know how to evaluate their specialty/subspecialty program and they would love a "how-to" guide on appropriate steps to do program evaluation. A few program directors accessed a program evaluation expert who was supporting all the PGME programs. However, if program evaluation is to be of value, then

one program evaluation expert for more than 60 programs, depending on the university, is not a realistic level of support. Additionally, program evaluators in my study noted that they are not seeing requests for proposals on the Canadian Evaluation Society or the AEA websites. These program evaluators explained that it is not their individual responsibility to advocate for program evaluation in PGME. Instead, they indicated that the Canadian Evaluation Society, as a professional association, is responsible for better communicating the advantages of evaluation planning and activities with external supports with the PGME departments. Finally, program directors who were attempting to find time in their busy schedules to do program evaluation noted that they distributed evaluation findings to faculty, residents, and PGME administration. The program evaluators added that is useful to include them in the interpretation and distribution of evaluation findings during "data parties" (E11), where everyone involved in the evaluation meets to support a thorough understanding of the findings.

There are strategies discussed in the literature about building capacities to do program evaluation. Preskill and Boyle (2008) and Cousins and Bourgeois (2014) noted that involvement in a program evaluation process (e.g., advisory committee, collaborative evaluation approaches) develops experiential learning from the planning and implementation of program evaluation, which establishes confidence to practice future evaluation. In my study, program directors suggested that they can develop their program evaluation abilities and expertise using a hands-on, guided method, and program evaluators noted that they can apply a participatory approach to encourage learning by doing to help CBD programs build their capacities to do program evaluation. Hotte et al. (2015) explained that by participating in program evaluations, individuals can gain access to ECB resources (i.e., staff and time), exposure to data collection and technology monitoring systems, and connections to evaluation templates, which they can apply towards maintaining ongoing and future evaluations. In this study, program directors insisted on the need to acquire program evaluation resources for guidance, and program evaluators noted that they can offer evaluation resources and expertise as guidance to build capacity and to save program directors' time. Hotte et al. (2015) encouraged including an evaluator to offer technical assistance and to form sustainable evaluation partnerships. To determine the dynamic of these evaluation partnerships, both program directors and program evaluators in my study advocated for clearer program evaluation policy expectations from the RCPSC, the PGME office, and within their specialties. My findings also identified another strategy for program evaluators to

build capacity to do program evaluation, which is to leverage existing data to simplify the program evaluation process.

***Capacities to Use Program Evaluation Findings***

In the literature, Bourgeois and Cousins (2013) reported that capacities to use evaluation processes and findings encourage evaluation literacy, informed organizational decision making, and learning benefits for stakeholders. In this study, evaluation literacy was intertwined with accreditation as program directors and program evaluators noted that the PGME department encouraged involvement in the evaluation process in preparation for or following accreditation. This involvement helped with organizational decision making, although program directors acknowledged using informal discussions and experiences as evaluation information. Moreover, program directors and program evaluators strongly suggested the importance of using findings that are connected to budget allocation and policy for program evaluation (i.e., at the RCPSC, PGME, and specialty/subspecialty levels). Program directors also noted that evaluation learning benefits can be the basis for action change, can alter stakeholders' understandings, and can encourage behavioural change from exposure to evaluation. However, program directors clarified that there were challenges to faculty buy-in to evaluation because even though faculty understood the value of evaluation, the reality of time constraints and the priorities of also working as practising physicians restricted the amount of time faculty could allocate to involvement in program evaluation.

Conceptual, instrumental, symbolic, and process use are different ways of using evaluation findings. Conceptual use is when evaluation findings are used to influence stakeholders' thinking about a program (Alkin & King, 2016). In this study, program directors acknowledged that the use of evaluation findings influenced faculty members' cognitive processing and how they perceived the program and its activities. For example, some program directors summarized the collected data and shared the findings in a presentation at their annual program retreat to spark discussion and to deepen faculty members' understanding of the program activities. Instrumental use is when program directors use evaluation findings to support decision-making or problem-solving processes (Shulha & Cousins, 1997). In this study, program directors shared that they applied evaluation findings to support direct actions, such as using the findings to advise the RCPSC of challenging EPAs to support making changes to the identified EPAs. Symbolic use leverages data to comply with reporting demands or to publicly support a

decision/opinion made on a different basis (Alkin & King, 2016). In this study, program directors and program evaluators indicated that they used collected data to meet the continual improvement standard when reporting to accreditation demands. Program evaluators also used standardized methodologies to support publication enticements as a way for program directors to publicly share their experiences with transitioning to CBD. Finally, process use is when participation in evaluation enhances use. The effects of process use include facilitating program directors' understandings of evaluation, supporting and reinforcing program intervention, facilitating CBD program development, infusing evaluation thinking into health care culture, and keeping program directors focused on program priorities (Patton, 2007). The program directors in this study agreed with Patton (2012) that building capacity to use evaluation findings encourages accountability between program resources, activities, outputs, and short- and long-term outcomes. However, the program directors were skeptical about whether the CBD transition will live up to the long-term outcomes of a better educational experience for residents leading to an improved Canadian health care system. Nevertheless, program evaluators emphasized how building capacity to use evaluation findings encourages social accountability between resources, activities, and short-term outcomes. The focus on short-term outcomes may be more feasible and, ultimately, may support the long-term outcome of ongoing improvement for residency programs.

Different strategies can encourage individuals to build their capacities to use program evaluation findings and improve CBD programs. Although evaluation findings usually support positive program improvement, the program directors in this study agreed with Cousins and Bourgeois (2014) that pushback to engage in program evaluation and use the findings stemmed from increased time commitments and no designated position for research and evaluation within their specialty/subspecialty programs. Cousins and Earl (1995) insisted that individuals conducting evaluations must demonstrate adequate technical rigour of methods to justify the findings as legitimate evidence to the users. For example, in this study, program evaluators leading a rapid cycle evaluation followed the Core Components Framework in order to generate valid and reliable findings for the program directors to use and publish. Furthermore, Patton (2015b) noted that the intended users should inform the evaluation decision-making process. In this study, program directors emphasized that to build capacity to use findings, program directors and faculty must contribute to the meaningful questions that are appropriate to the stage of the

program to establish buy-in for program evaluation to increase the likeliness of using the findings. Program evaluators also recommended that program evaluation is included at the beginning of program design and implementation, includes the stakeholders who will be using the findings, and focuses stakeholder questions on meaningful program areas over which they hold authority to make changes. In the literature, debriefing and sharing findings are listed as ways to hold individuals accountable and responsible for sharing their successes and challenges of implementing the findings with other stakeholders (Patton, 2008). In this study, program directors noted that to build capacity to use findings, they can share findings with faculty, residents, and other programs and they can also understand evaluation as helpful for the accreditation process in order to maintain accountability for implementing findings. Moreover, program evaluators encouraged the use of program evaluation approaches that are appropriate to the CBD implementation context, advocated for interpretation of the findings alongside the stakeholders to co-develop a manageable follow-up plan, and encouraged programs to share findings, which establishes an accountability cycle to make changes based on findings. Thus, this study adds the importance of program evaluators making sense of findings alongside program directors to avoid findings being "shelved" and to scaffold an action implementation cycle that verifies the extent to which findings are being applied.

**Contributions of the Findings**

The findings from my study contribute to the literature on how to do quality assurance on CBD, practice in program evaluation, and methods pertaining to program evaluation in a variety of ways. First, they contribute to quality assurance by extending our knowledge of program evaluation in CBD. Despite increased demands for program evaluation, prior to this study, the medical education community in Canada knew very little about the ongoing evaluation activities in CBD. However, this paucity of empirical evidence concerning evaluation is not restricted to the medical education field. For many years, researchers have been calling for more and better-quality research on evaluation (Cousins & Earl, 1999; Henry & Mark, 2003; Mark, 2008; Smith, 1993). Through my exploration of evaluation practices in this specific context, I have demonstrated where improvements are needed to improve practice in Canadian CBD programs.

Second, in terms of practice, my findings assist with guiding future evaluation activities and contribute to researchers' and evaluators' understandings of important evaluation issues, including collaborating with interested stakeholders, leading an evaluation approach that is

manageable, using methods that complement the environment, and offering professional development sessions about program evaluation. My findings also advise on practical strategies for future evaluation activities to build capacities to do program evaluation (e.g., apply a participatory approach, offer evaluation resources and expertise, leverage existing data, advocate for clear program evaluation policy expectations). Additionally, regarding practice, my findings provide guidance on activities that build capacities to use evaluation findings (e.g., recommend that program evaluation is included at the beginning, include the stakeholders who will be using the findings, encourage the use of appropriate program evaluation approaches, and interpret the findings alongside the stakeholders; Mark, 2008). Furthermore, by documenting and understanding the current state of program evaluation in this field, my findings establish what types of resources (e.g., time, money, human resources, improved technology infrastructure, faculty buy-in) and activities (e.g., advocating for clearer program evaluation expectations from the RCPSC, asking meaningful questions that are appropriate to the stage of the program, building knowledge about appropriate evaluation methods, and sharing findings) are needed to advance program evaluation in CBD.

Lastly, in terms of contributions to methods, in this study, I followed a three-phase, explanatory sequential research design (Creswell & Creswell, 2018). Phase 1 focused on the collection and analysis of quantitative survey data, Phase 2 involved the collection and analysis of qualitative interview data, and Phase 3 included qualitative interview data from a different population. I collected and analyzed both the quantitative and qualitative data separately. However, the Phase 1 survey findings informed Phase 2's interview guide, the Phase 2 interview findings informed Phase 3's interview guide, and I triangulated these findings in this chapter (Creswell & Plano Clark, 2018). My three-phase design extended the breadth and depth of inquiry into the use of program evaluation in CBD and how to build capacity to do evaluation and use evaluation findings within this context. This study design may encourage other medical education researchers to explore other topics in order to answer complex research questions that are well-suited to mixed methods. By keeping the focus on the research questions and enabling the gathering and integration of multiple types of data, they can realize the advantages of pragmatism and reach interesting and insightful conclusions.

# Chapter 8: Conclusion

In this chapter, I discus the study's limitations. The chapter concludes with my suggestions for future research in this area and my concluding remarks.

## Limitations of the Study

All research designs have both strengths and limitations (Fraenkel et al., 2019). In the following subsections, I discuss the different limitations of each phase of my study in detail.

### *Limitations of Phase 1*

I used an internet-based survey in Phase 1, which is associated with the advantages of convenience, minimal costs, faster turnaround, multimedia interface, use of portable devices, and reduced data entry (Fraenkel et al., 2019). However, internet-based surveys also include disadvantages, such as lower response rates and the potential for erroneous data due to quick data entry facilitated by computers (Fraenkel et al., 2019). Regarding the response rates, I reviewed the RCPSC website to determine which specialties/subspecialties had transitioned to CBD. Of the 445 program directors that I identified and invited to participate, 149 completed the survey. Thus, the response rate for Phase 1 was 33.5%, which is equivalent or higher than other response rates in journal articles that surveyed program directors in graduate medical education. For example, Aguwa et al. (2022) examined underrepresented minority recruitment and diversity education initiatives in by residency programs and surveyed residency program leadership (i.e., residency program directors and associate residency program directors); 63 of 188 participants responded (33.5% response rate). My survey response rate is also higher than Al Achkar et al. (2018), who examined current trends in program directors in interprofessional education within American graduate medical education, and whose survey was completed by 233 of 1757 program directors (13.3% response rate).

One limitation of my internet-based survey is coverage error, which occurs when the list from which the sample members was acquired does not accurately represent the population that the survey data is intended to represent (Dillman et al., 2014). Coverage error is the gap between the findings produced from an inaccurate list and what would have been discovered with an accurate list (Dillman et al., 2014). I acquired the contact emails of the program directors from the RCPSC website, and I tracked email delivery, which indicated that 373 (83.8%) of the emails were successfully delivered. Thus, my sample list included inaccurate emails, as listed on the RCPSC website. Finally, although programs were supposed to transition to CBD during specific

years, due to COVID-19 or other logistical reasons, some of the programs had not yet transitioned and it is possible that I omitted other programs that transitioned early from my sample population.

Another limitation of Phase 1 of my study is sampling error, which is the difference between the findings generated when only a sample of the population is surveyed compared to the findings when every member of the population is surveyed (Dillman et al., 2014). Sampling error is present because I decided to survey only some program directors by sending only three reminder emails to complete the survey. Had I continued to try to sample every person on the list with more reminder emails, I may have acquired slightly different findings.

Additionally, I encountered a nonresponse error during Phase 1, which is the difference between the findings when only some of the sampled population responds compared to when all of the population responds (Dillman et al., 2014). This limitation occurs when the population who does not respond differs from the sample that do in a way that skews the findings (Dillman et al., 2014). The program directors who completed my survey noted the challenges of time, money, human resources, and technology infrastructure as barriers to doing program evaluation, which I assume applies to most PGME programs. Nevertheless, the program directors who responded to my survey are most likely those who have an interest in program evaluation, and if they find it challenging to do program evaluation, then the program directors who did not respond to my survey are even less likely to be engaging in program evaluation. Thus, I believe that my findings are representative of the minimal program evaluation that is actually occurring in PGME programs.

Furthermore, measurement error is another possible limitation of my study. Measurement error is the difference between the findings and the actual value because respondents provided inaccurate answers to the survey questions (Dillman et al., 2014). This error occurs when the survey respondents are not able or willing to offer accurate responses due to poor design of the questions or survey mode effects (i.e., an online survey may encourage respondents to be more honest, but the absence of an interviewer for motivation and assistance can reduce the data quality; Dillman et al., 2014).

Another potential survey limitation is the possibility that program directors self-excluded from the survey due to the email subject line (i.e., self-selection bias; Cook & Campbell, 1979). In the open-ended survey responses, some program directors indicated that they were unclear

about what I meant by program evaluation. Other program directors who did not complete the survey could have decided the survey was not relevant to them due to their lack of understanding of the term "program evaluation" in the email subject line. If those program directors had completed the survey, it would have generated a higher response rate and provided a better understanding of the extent of CBD programs engagement in program evaluation.

Finally, respondents' survey answers may have been limited because of their desire to present socially acceptable answers or because the survey included mostly closed-ended questions (Cook & Campbell, 1979). I used closed-ended questions to make data collection more straightforward and to facilitate data analysis. However, the closed-ended format may have posed a survey instrumentation threat by limiting the acquisition of respondents' complete perceptions (Cook & Campbell, 1979). Despite these possible limitations, the survey results offered me insight about how to shape my Phase 2 follow-up questions.

***Limitations of Phase 2***

The Phase 2 interviews relied on program directors' reports of their perceptions. Due to my presence in the interviews, participants may have offered socially desirable responses to reflect themselves and their interests more favourably. It is also possible that the program directors who agreed to be interviewed reflect a particular group of program directors and do not accurately represent program directors in general, thereby resulting in selection bias (Cook & Campbell, 1979). As such, there are suspected differences between the program directors who participated in Phase 2 and those who did not. It is likely that program directors who participated in this phase were more outspoken, active in their professions as program directors, and interested in the topic area of program evaluation in medical education. Thus, these particular program directors may have had different perceptions than the program directors who did not participate in my study. Nevertheless, these interview responses supported my development of the Phase 3 program evaluator guide and interview protocols.

***Limitations of Phase 3***

I acquired email addresses for certain Canadian English-speaking program evaluators from the AEA website. Based on the relevance of their designated affiliation, I contacted 33 eligible members from two TIGs: Health Evaluation and Health Professions Education, Evaluation and Research. In addition to these 33 evaluators, I contacted 4 program evaluators who were known to evaluate medical education programs in Canada but were not AEA TIG

members. I also contacted 15 program evaluators associated with Canadian English PGME programs. In total, 12 of the 52 program evaluators (23.1% response rate) that I contacted agreed to participate in an interview. This response rate may be viewed as a limitation because it is plausible that data themes would have generated different results if more program evaluators had accepted an interview. However, it is comparable to the response rate for a survey targeted at program evaluators whose emails were acquired from the AEA TIG listings. Wanzer (2020) surveyed 233 AEA program evaluators out of 985 possible program evaluators (23.7% response rate) to understand how evaluation is defined and asked how, if at all, they differentiated evaluation from research. Thus, "despite research suggesting that AEA members consider research on evaluation as important, response rates for research on evaluation studies are often only between 10–30%" (Wanzer, 2018). Therefore, my interview response rate from program evaluators is in line with the survey response rates of program evaluators reported in the literature.

Another limitation related to the Phase 3 interview findings is the potential that program evaluators self-excluded from the interview due to the email subject line in my initial request (i.e., self-selection bias; Cook & Campbell, 1979). Some program evaluators informed me via an email response that they had decided the interview was not relevant for them due to the term "Program Evaluation in CBME Interview Request" in the title. These program evaluators mentioned that they did not feel that they had experience to share about evaluating CBME programs, but indicated that they had done evaluations in medical education. After following up with these individuals, some still refused the interview, but their experience in evaluating medical education more generally could have contributed to more informed findings.

Finally, the Phase 3 interviews relied on program evaluators' reports of their perceptions. Participants may have provided socially desirable responses to reflect themselves and their interests more favourably due to my presence in the interviews. It is also possible that the program evaluators who agreed to be interviewed reflect a particular group of program evaluators and do not accurately represent program evaluators in general, thus resulting in selection bias (Cook & Campbell, 1979). As such, there may be differences between the program evaluators who participated and those who did not. It is likely that the program evaluators who participated in this phase were more forthright, active in their professions as program evaluators, and interested in the topic area of program evaluation in medical education. Thus, these

particular program evaluators may have highlighted different perceptions than the program evaluators who did not participate in my study. Nevertheless, these interview responses offered insight into how program evaluators are currently or could be supporting CBD programs in program evaluation and how program evaluators can help CBD programs build their capacities to do program evaluation and use evaluation findings.

**Future Research**

In terms of future research, I recommend that researchers who are interested in program evaluation apply a quality assurance frame of mind to look at programs that were successful in transitioning to CBD compared to programs that are having issues. Researchers should study factors that contribute to the success or failure as a next step. I anticipate that programs with a more defined program evaluation plan would be adapting better to the CBD transition. . The RCPSC program evaluation dashboard (RCPSC, 2021) is currently collecting data about conversations on CBD program evaluation. One future area of focus is to explore the extent to which the findings from these program evaluations are being used. Specifically, findings from the RCPSC National Program Evaluation Studies identify three challenging tasks: conducting faculty development, engaging off-service disciplines, and enlisting adequate support staff (RCPSC, 2021). These findings directly relate to and impact program evaluation efforts and it would be interesting to further explore the implications of these findings on program evaluation in the various CBD programs. Another potential future research question is how are program directors being supported so they are able to evaluate their CBD programs? Additionally, future research could investigate successful strategies for offering professional development sessions about program evaluation to all CBD stakeholders by asking, for example, how can we engage CBD stakeholders in professional development about program evaluation? Another next step would be to target the non-responders to determine what is going on within their program. I suspect some program directors are overwhelmed with having to do program evaluation and they are doing the bare mreainimum to implement program evaluation. However, the bare minimum would not include program evaluation.

With more time, outcome-focused evaluation approaches could be applied to determine the extent to which the transition to CBD has prepared residents, thereby providing better health care for patients. I also recommend that future research include other stakeholders (i.e., RCPSC administration, PGME leadership, other competence committee faculty, and residents) to learn

more about the creation of an evaluation culture and mindset in CBD. Future research could investigate how to provide resources to build capacity to do program evaluation (e.g., templates shared across Canada or specialty/subspecialty tools and resources). Specifically, the research question could ask which program evaluation tools and resources are most beneficial for CBD?

**Conclusion**

My thesis explored program evaluation in medical education. Through this study, I sought to understand the extent to which CBD programs are engaging in program evaluation, the reasons why CBD programs are engaging or not engaging in program evaluation, and the ways in which CBD programs can build their capacities to do program evaluation and use evaluation findings. I also investigated how program evaluators are currently supporting CBD programs in program evaluation and how program evaluators can help CBD programs build their capacities to do program evaluation and use evaluation findings. Using mixed methods, including surveys and interviews, I examined the perceptions of program directors and program evaluators on this topic. Overall, the findings show that CBD programs are struggling to engage in program evaluation because of limited available resources (i.e., cost, time, and expertise constraints, and lack of technology infrastructure) and buy-in. Nevertheless, when they do engage in program evaluation, CBD programs are using ad hoc evaluation methods and a team-based format. Program directors can develop expertise in their program evaluation abilities, acquire program evaluation resources for guidance, and advocate for clear program evaluation expectations from the RCPSC, the PGME office, and within their specialties. Program directors can also ask meaningful questions that are appropriate to the stage of their program, establish buy-in for program evaluation to increase the likeliness of using the findings, share findings, and understand evaluation as helpful for the accreditation process.

My findings also revealed that program evaluators are supporting CBD programs in program evaluation by responding in a reactive way as a temporary and external evaluation consultants, collaborating with interested stakeholders as internal evaluation experts from a university, and leading an evaluation approach that is manageable for program directors with limited time. Program evaluators recommend using methods that complement the PGME environment, using evaluation processes that lead to the publication of findings, and offering professional development sessions about program evaluation to support accreditation. Furthermore, program evaluators can help CBD programs build their capacities to do program

evaluation by using a participatory evaluation approach, offering evaluation resources and expertise as guidance, leveraging existing data, and advocating for clearer program evaluation policy expectations from the RCPSC. Finally, program evaluators can help CBD programs build their capacities to use evaluation findings by recommending that it is included at the beginning of program design and implementation, including the stakeholders who will be using the findings, and focusing stakeholder questions on meaningful program areas over which they hold authority to make changes. Program evaluators can also encourage the use of program evaluation approaches that are appropriate to the CBD implementation context, interpret the findings alongside stakeholders and co-develop a manageable follow-up plan, and encourage programs to share findings, which establishes an accountability cycle.

This study contributes further empirical research to confirm and expand on the discussion about program evaluation in medical education. I expect that these findings will support program evaluators in engaging in program evaluation in medical education. This work is important because it can support the ongoing transition to CBD in PGME departments to develop better prepared residents, which will, in turn, improve health care for patients. Developing this program evaluation mindset is a culture shift that offers life-changing opportunities for programs, evaluators, residents, and communities.

## References

Aguwa, U. T., Wang, J., Woreta, F., Pettey, J., Sun, G., Pineles, S., & Srikumaran, D. (2022). Residency program diversity recruitment and education: Survey of efforts and barriers to implementation. *Journal of Surgical Education*, *79*(3), 595–605. https://doi.org/10.1016/j.jsurg.2021.11.016

Al Achkar, M., Hanauer, M., Colavecchia, C., & Seehusen, D. A. (2018). Interprofessional education in graduate medical education: Survey study of residency program directors. *BMC Medical Education*, *18*(1), Article 11. https://doi.org/10.1186/s12909-017-1104-z

Alkin, M. C., & King, J. A. (2016). The historical development of evaluation use. *American Journal of Evaluation*, *37*(4), 568–579. https://doi.org/10.1177/1098214016665164

Alkin, M. C., & Christie, C. (2012). An evaluation theory tree. In M. C. Alkin (Ed.), *Evaluation roots: A wider perspective of theorists' views and influences* (2nd ed., pp. 13–65). SAGE Publications. https://doi.org/10.4135/9781412984157

Bourgeois, I., & Cousins, J. B. (2013). Understanding dimensions of organizational evaluation capacity. *American Journal of Evaluation*, *34*(3), 299–319. https://doi.org/10.1177/1098214013477235

Canadian Evaluation Society. (2018). *Competencies for Canadian evaluation practice*. https://evaluationcanada.ca/txt/2_competencies_cdn_evaluation_practice_2018.pdf

CanRAC. (2020). *General standards of accreditation for residency programs*. https://www.royalcollege.ca/rcsite/documents/canera/general-standards-accreditation-for-residency-programs-e

Clarke, A. (2006). Evidence-based evaluation in different professional domains: Similarities, differences and challenges. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The SAGE handbook of evaluation* (pp. 559–581). SAGE Publications. https://doi.org/10.4135/9781848608078.n25

Compton, D. W., MacDonald, G., Baizerman, M., Schooley, M., & Zhang, L. (2008). Using evaluation capacity building (ECB) to interpret evaluation strategy and practice in the United States National Tobacco Control Program (NTCP): A preliminary study. *Canadian Journal of Program Evaluation*, *23*(3), 199–224. https://evaluationcanada.ca/system/files/cjpe-entries/23-3-199.pdf

Cook, D. (2010). Twelve tips for evaluating educational programs. *Medical Teacher*, *32*(4), 296–301. https://doi.org/10.3109/01421590903480121

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Houghton Mifflin.

Cousins, J. B. (2005). Will the real empowerment evaluation please stand up?: A critical friend perspective. In D. M. Fetterman & A. Wandersman (Eds.), *Empowerment evaluation principles in practice* (pp. 183–208). The Guilford Press.

Cousins, J. B., & Bourgeois, I. (2014). Multiple case study methods and findings. *New Directions for Evaluation*, *2014*(141), 25–99. https://doi.org/10.1002/ev.20077

Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, *14*(4), 397–418. https://doi.org/10.3102/01623737014004397

Cousins, J. B., & Earl, L. M. (1995). The case for participatory evaluation: Theory, research, practice. In *Participatory evaluation in education: Studies of evaluation use and organizational learning* (pp. 3–20). Falmer Press.

Cousins, J. B., & Earl, L. M. (1999). When the boat gets missed: Response to M. F. Smith. *American Journal of Evaluation*, *20*(2), 309–317. https://doi.org/10.1177/109821409902000212

Cousins, J. B., Elliott, C., Amo, C., Bourgeois, I., Chouinard, J., Goh, S. C., & Lahey, R. (2008). Organizational capacity to do and use evaluation: Results of a pan-Canadian survey of evaluators. *Canadian Journal of Program Evaluation*, *23*(3), 1–35. https://evaluationcanada.ca/system/files/cjpe-entries/23-3-001.pdf

Cousins, J. B., Goh, S. C., Clark, S., & Lee, L. E. (2004). Integrating evaluative inquiry into the organizational culture: A review and synthesis of the knowledge base. *Canadian Journal of Program Evaluation*, *19*(2), 99–141. https://evaluationcanada.ca/secure/19-2-099.pdf

Cousins, J. B., Hay, K., & Chouinard, J. (2015). The third perspective: Uniting accountability and learning within an evaluation framework that takes a moral-political stance. In C. A. Christie & A. T. Vo (Eds.), *Evaluation use and decision making in society: A tribute to Marvin C. Alkin* (pp. 91–112). Information Age Publishing, Inc.

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, *1998*(80), 5–23. https://doi.org/10.1002/ev.1114

Creswell, J., & Creswell, J. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Creswell, J., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Creswell, J. W. (2014). *A concise introduction to mixed methods research*. SAGE Publications.

Creswell, J. W. (2015). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (5th ed.). Pearson.

Creswell, J. W., Klassen, A. C., Plano Clark, V. L., & Smith, K. C. (2011). *Best practices for mixed methods research in the health sciences*. National Institutes of Health. https://obssr.od.nih.gov/sites/obssr/files/Best_Practices_for_Mixed_Methods_Research.pdf

Davis, B. (2004). *Inventions of teaching: A genealogy*. Lawrence Erlbaum Associates.

Dewey, J. (1998). The development of American pragmatism. In L. A. Hickman & T. M. Alexander (Eds.), *The essential Dewey: Vol 1. Pragmatism, education, democracy*. (Original work published in 1925). Indiana University Press.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). John Wiley & Sons, Inc.

Doll, W. E., Jr., & Trueit, D. (2010). Complexity and the health care professions. *Journal of Evaluation in Clinical Practice*, *16*(4), 841–848. https://doi.org/10.1111/j.1365-2753.2010.01497.x

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education* (10th ed.). McGraw Hill Education.

Frank, J. R., Mungroo, R., Ahmad, Y., Wang, M., De Rossi, S., & Horsley, T. (2010). Toward a definition of competency-based education in medicine: A systematic review of published definitions. *Medical Teacher*, *32*(8), 631–637. https://doi.org/10.3109/0142159X.2010.500898

Frye, A. W., & Hemmer, P. A. (2012). Program evaluation models and related theories: AMEE Guide No. 67. *Medical Teacher*, *34*(5), 288–299. https://doi.org/10.3109/0142159X.2012.668637

Funnell, S. C., & Rogers, P. J. (2011). The essence of program theory. In *Purposeful program theory: Effective use of theories of change and logic models* (pp. 3–14). John Wiley & Sons, Incorporated.

Giddings, L. S., & Grant, B. (2007). A trojan horse for positivism?: A critique of mixed methods research. *Advances in Nursing Science*, *30*(1), 52–60. https://doi.org/10.1097/00012272-200701000-00006

Goldie, J. (2006). AMEE Education Guide no. 29: Evaluating educational programmes. *Medical Teacher*, *28*(3), 210–224. https://doi.org/10.1080/01421590500271282

Greene, J. C. (2006). Toward a methodology of mixed methods social inquiry. *Research in the Schools*, *13*(1), 93–98. https://researchgate.net/publication/228968099_Toward_a_methodology_of_mixed_methods_social_inquiry

Greene, J. C., & Caracelli, V. J. (1997). *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms*. Jossey-Bass Publishers.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, *11*(3), 255–274. https://doi.org/10.3102/01623737011003255

Gruppen, L. D., ten Cate, O., Lingard, L. A., Teunissen, P. W., & Kogan, J. R. (2018). Enhanced requirements for assessment in a competency-cased, time-variable medical education system. *Academic Medicine*, *93*(3S), S17–S21. https://doi.org/10.1097/ACM.0000000000002066

Hall, A. K., Rich, J., Dagnone, J. D., Weersink, K., Caudle, J., Sherbino, J., Frank, J. R., Bandiera, G., & Van Melle, E. (2020). It's a marathon, not a sprint: Rapid evaluation of competency-based medical education program implementation. *Academic Medicine*, *95*(5), 786–793. https://doi.org/10.1097/ACM.0000000000003040

Harnar, M. A., & Preskill, H. (2007). Evaluators' descriptions of process use: An exploratory study. *New Directions for Evaluation*, *2007*(116), 27–44. https://doi.org/10.1002/ev.241

Henry, G. T., & Mark, M. M. (2003). Toward an agenda for research on evaluation. *New Directions for Evaluation*, *2003*(97), 69–80. https://doi.org/10.1002/ev.77

Holmboe, E. S., Ward, D. S., Reznick, R. K., Katsufrakis, P. J., Leslie, K. M., Patel, V. L., Ray, D. D., & Nelson, E. A. (2011). Faculty development in assessment: The missing link in competency-based medical education. *Academic Medicine*, *86*(4), 460–467. https://doi.org/10.1097/ACM.0b013e31820cb2a7

Horsley, T., Moreau, K., Lockyer, J., Zeiter, J., Varpio, L., & Campbell, C. (2016). More than reducing complexity: Canadian specialists' views of the Royal College's maintenance of certification framework and program. *Journal of Continuing Education in the Health Professions*, *36*(3), 157–163. https://doi.org/10.1097/CEH.0000000000000099

Hotte, N., Simmons, L., Beaton, K., & LDCP Workgroup. (2015). *Scoping review of evaluation capacity building strategies*. Public Health Ontario.

Iobst, W. F., Sherbino, J., Cate, ten Cate, O., Richardson, D. L., Dath, D., Swing, S. R., Harris, P., Mungroo, R., Holmboe, E. S., & Frank, J. R. (2010). Competency-based medical education in postgraduate medical education. *Medical Teacher*, *32*(8), 651–656. https://doi.org/10.3109/0142159X.2010.500709

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, *1*(2), 112–133. https://doi.org/10.1177/1558689806298224

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26. https://doi.org/10.3102/0013189X033007014

King, J. A. (2007). Developing evaluation capacity through process use. *New Directions for Evaluation*, *2007*(116), 45–59. https://doi.org/10.1002/ev.242

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of the American Society of Training Directors*, *13*, 3–9.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development* (pp. 301–319). McGraw Hill.

Kirkpatrick, D. L. (1996). *Evaluating training programs: The four levels* (1st ed.). Berrett-Koehler.

Kirkpatrick, D. L. (1998). *Evaluating training programs: The four levels* (2nd ed.). Berrett-Koehler.

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

Kirkpatrick Partners. (2020). *The new world Kirkpatrick model*. https://www.kirkpatrickpartners.com/Our-Philosophy/The-New-World-Kirkpatrick-Model

Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*, *33*(3), 307–338. https://doi.org/10.1177/1098214011434608

Lancaster, G. A., Dodd, S., & Williamson, P. R. (2004). Design and analysis of pilot studies: Recommendations for good practice. *Journal of Evaluation in Clinical Practice*, *10*(2), 307–312. https://doi.org/10.1111/j..2002.384

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.

Lincoln, Y. S., & Guba, E. G. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The handbook of qualitative research* (2nd ed., pp. 1065–1122). SAGE Publications.

Lune, H., & Berg, B. L. (2017). *Qualitative research methods for the social sciences* (9th ed.). Pearson.

Mark, M. M. (2008). Building a better evidence base for evaluation theory: Beyond general calls to a framework of types of research on evaluation. In N. L. Smith & P. R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 111–134). Guilford Press.

Mayne, J. (2011). Contribution analysis: Addressing cause and effect. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution and beyond*. Transaction Publishers.

Mayne, J. (2019). Revisiting contribution analysis. *Canadian Journal of Program Evaluation*, *34*(2), 171–191. https://doi.org/10.3138/cjpe.68004

Mertens, D. M., & Wilson, A. T. (2012). *Program evaluation theory and practice: A comprehensive guide*. The Guilford Press.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2019). *Qualitative data analysis: A methods sourcebook* (4th ed.). SAGE Publications.

Millar, M. M., & Dillman, D. A. (2011). Improving response to web and mixed-mode surveys. *Public Opinion Quarterly*, *75*(2), 249–269. https://doi.org/10.1093/poq/nfr003

Moreau, K. A. (2017). Twelve tips for planning and conducting a participatory evaluation. *Medical Teacher*, *39*(4), 334–340. https://doi.org/10.1080/0142159X.2017.1286310

Moreau, K. A., & Eady, K. (2015). Connecting medical education to patient outcomes: The promise of contribution analysis. *Medical Teacher*, *37*(11), 1060–1062. https://doi.org/10.3109/0142159X.2015.1060307

Naccarella, L., Pirkis, J., Kohn, F., Morley, B., Burgess, P., & Blashki, G. (2007). Building evaluation capacity: Definitional and practical implications from an Australian case study. *Evaluation and Program Planning*, *30*(3), 231–236. https://doi.org/10.1016/j.evalprogplan.2007.05.001

Oandasan, I., Martin, L., McGuire, M., & Zorzi, R. (2020). Twelve tips for improvement-oriented evaluation of competency-based medical education. *Medical Teacher*, *42*(3), 272–277. https://doi.org/10.1080/0142159X.2018.1552783

Parsons, B. L., Lovato, C. Y., Hutchinson, K., & Wilson, D. (2016). Building evaluation capacity through CLIPs: Communities of learning, inquiry, and practice. *New Directions for Evaluation, 2016*(151), 21–36. https://doi.org/10.1002/ev.20199

Patton, M. Q. (1988). Paradigms and pragmatism. In D. M. Fetterman (Ed.), *Qualitative approaches to evaluation in education: The silent scientific revolution* (pp. 116–137). Praeger Publishers.

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

Patton, M. Q. (2007). Process use as a usefulism. *New Directions for Evaluation, 2007*(116), 99–112. https://doi.org/10.1002/ev.246

Patton, M. Q. (2008). Fostering intended use by intended users: The personal factor. In *Utilization-focused evaluation* (4th ed., p. 667). SAGE Publications.

Patton, M. Q. (2012). A utilization-focused approach to contribution analysis. *Evaluation*, *18*(3), 364–377. https://doi.org/10.1177/1356389012449523

Patton, M. Q. (2015a). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). SAGE Publications.

Patton, M. Q. (2015b). The sociological roots of utilization-focused evaluation. *The American Sociologist, 46*(4), 457–462. https://doi.org/10.1007/s12108-015-9275-8

Preskill, H. (2008). Evaluation's second act: A spotlight on learning. *American Journal of Evaluation, 29*(2), 127–138. https://doi.org/10.1177/1098214008316896

Preskill, H., & Boyle, S. (2008). Insights into evaluation capacity building: Motivations, strategies, outcomes, and lessons learned. *Canadian Journal of Program Evaluation*, *23*(3), 147–174. https://evaluationcanada.ca/system/files/cjpe-entries/23-3-147.pdf

Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2019). *Evaluation: A systematic approach*. SAGE Publications.

Royal College of Physicians and Surgeons of Canada. (2019). *CBD program evaluation recommendations report: Findings from FMRQ, RCPSC*. https://www.royalcollege.ca/rcsite/documents/cbd/recommendations-report-2019-e.pdf

Royal College of Physicians and Surgeons of Canada. (2020a). *CBD start, launch and exam schedule*. http://www.royalcollege.ca/rcsite/documents/cbd-community-touchpoint/cbd-rollout-schedule-e

Royal College of Physicians and Surgeons of Canada. (2020b). *Competence by Design*. http://www.royalcollege.ca/rcsite/cbd/what-is-cbd-e

Royal College of Physicians and Surgeons of Canada. (2020c). *Competence by Design Launch Schedule*. http://www.royalcollege.ca/rcsite/cbd/cbd-implementation-e

Royal College of Physicians and Surgeons of Canada. (2021). *Competence by Design – Program evaluation dashboard*. https://app.powerbi.com/view?r=eyJrIjoiMjI2M2E1NTUtZGUxOS00NjQwLWJlYTctNDYyYjRiMmVkZmIzIiwidCI6ImE2YTgzYWYwLTcwODktNDA5MC05NDI1LTM3Mzg2MzhiNTFlNSJ9

Royal College of Physicians and Surgeons of Canada. (2022). *What is Competence by Design?* https://www.royalcollege.ca/rcsite/cbd/what-is-cbd-e

Seidman, I. (2019). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (5th ed.). Teachers College Press.

Shah, N., Desai, C., Jorwekar, G., Badyal, D., & Singh, T. (2016). Competency-based medical education: An overview and application in pharmacology. *Indian Journal of Pharmacology*, *48*(7), 5–9. https://doi.org/10.4103/0253-7613.193312

Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research, and practice since 1986. *American Journal of Evaluation, 18*(3), 195–208. https://doi.org/10.1177/109821409701800302

Smith, N. L. (1993). Improving evaluation theory through the empirical study of evaluation practice. *American Journal of Evaluation, 14*(3), 237–242. https://doi.org/10.1177/109821409301400302

Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation*, *2002*(93), 7–26. https://doi.org/10.1002/ev.39

Stufflebeam, D., & Shinkfield, A. (2007). *Evaluation theory, models, & applications*. Jossey Bass/John Wiley & Sons, Inc.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining the qualitative and quantitative approaches*. SAGE Publications.

Tashakkori, A., & Teddlie, C. (2003). The past and future of mixed methods research: From data triangulation to mixed model designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 671–702). SAGE Publications.

Torres, R. T., Preskill, H. S., & Piontek, M. E. (2005). *Evaluation strategies for communicating and reporting: Enhancing learning in organizations* (2nd ed.). SAGE Publications.

Van Melle, E., Frank, J., Brzeznia, S., & Gorman, L. (2017). *Competence by design–Residency education: A framework for program evaluation*. Royal College of Physicians and Surgeons of Canada. http://www.royalcollege.ca/rcsite/documents/awards-grants/cbd-re-pe-november-2017.pdf

Van Melle, E., Frank, J. R., Holmboe, E. S., Dagnone, D., Stockley, D., & Sherbino, J. (2019). A core components framework for evaluating implementation of competency-based medical education programs. *Academic Medicine*, *94*(7), 1002–1009. https://doi.org/10.1097/ACM.0000000000002743

Van Melle, E., Gruppen, L., Holmboe, E. S., Flynn, L., Oandasan, I., & Frank, J. R. (2017). Using contribution analysis to evaluate competency-based medical education programs: It's all about rigor in thinking. *Academic Medicine*, *92*(6), 752–758. https://doi.org/10.1097/ACM.0000000000001479

Vassar, M., Wheeler, D. L., Davison, M., & Franklin, J. (2010). Program evaluation in medical education: An overview of the utilization-focused approach. *Journal of Educational Evaluation for Health Professions, 7*(1), 1–3. https://doi.org/10.3352/jeehp.2010.7.1

Wanzer, D. L. (2018, April 3). Research on evaluation: It takes a village (the problem). *DanaWanzer*. https://danawanzer.com/roe-theproblem/

Wanzer, D. L. (2020). What is evaluation?: Perspectives of how evaluation differs (or not) from research. *American Journal of Evaluation*, *42*(1), 28–46. https://doi.org/10.1177/1098214020920710

## Appendix A: University of Ottawa Social Sciences and Humanities Research Ethics Board Approval

27/10/2021

Université d'Ottawa
Bureau d'éthique et d'intégrité de la recherche

University of Ottawa
Office of Research Ethics and Integrity

**CERTIFICAT D'APPROBATION ÉTHIQUE | CERTIFICATE OF ETHICS APPROVAL**

| | |
|---|---|
| **Numéro du dossier / Ethics File Number** | S-01-21-6537 |
| **Titre du projet / Project Title** | Evaluation in Competency-Based Medical Education: A Mixed Methods Study |
| **Type de projet / Project Type** | Thèse de doctorat / Doctoral thesis |
| **Statut du projet / Project Status** | Approuvé / Approved |
| **Date d'approbation (jj/mm/aaaa) / Approval Date (dd/mm/yyyy)** | 22/02/2021 |
| **Date d'expiration (jj/mm/aaaa) / Expiry Date (dd/mm/yyyy)** | 21/02/2022 |

**Équipe de recherche / Research Team**

| Chercheur / Researcher | Affiliation | Role |
|---|---|---|
| Jenna MILOSEK | Faculté d'éducation / Faculty of Education | Chercheur Principal / Principal Investigator |
| Katherine MOREAU | Faculté d'éducation / Faculty of Education | Superviseur / Supervisor |

**Conditions spéciales ou commentaires / Special conditions or comments**

Feb 2021: Only Phase 1 (survey) of the study was approved.

Jun 2021: Only Phase 1 (survey) and Phase 2 (interviews with Program Directors) of the study are approved.

Oct 2021: Phases 1, 2, and 3 are approved.

550, rue Cumberland, pièce 154 550 Cumberland Street, Room 154
Ottawa (Ontario) K1N 6N5 Canada Ottawa, Ontario K1N 6N5 Canada

613-562-5387 • 613-562-5338 • ethique@uOttawa.ca / ethics@uOttawa.ca
www.recherche.uottawa.ca/deontologie | www.recherche.uottawa.ca/ethics

27/10/2021

Université d'Ottawa

Bureau d'éthique et d'intégrité de la recherche

University of Ottawa

Office of Research Ethics and Integrity

Le Comité d'éthique de la recherche (CÉR) de l'Université d'Ottawa, opérant conformément à l'*Énoncé de politique des Trois conseils* (2014) et toutes autres lois et tous règlements applicables, a examiné et approuvé la demande d'éthique du projet de recherche ci-nommé.

L'approbation est valide pour la durée indiquée plus haut et est sujette aux conditions énumérées dans la section intitulée "Conditions Spéciales ou Commentaires". Le formulaire « Renouvellement ou Fermeture de Projet » doit être complété quatre semaines avant la date d'échéance indiquée ci-haut afin de demander un renouvellement de cette approbation éthique ou afin de fermer le dossier.

Toutes modifications apportées au projet doivent être approuvées par le CÉR avant leur mise en place, sauf si le participant doit être retiré en raison d'un danger immédiat ou s'il s'agit d'un changement ayant trait à des éléments administratifs ou logistiques du projet. Les chercheurs doivent aviser le CÉR dans les plus brefs délais de tout changement pouvant augmenter le niveau de risque aux participants ou pouvant affecter considérablement le déroulement du projet, rapporter tout événement imprévu ou indésirable et soumettre toute nouvelle information pouvant nuire à la conduite du projet ou à la sécurité des participants.

The University of Ottawa Research Ethics Board, which operates in accordance with the *Tri-Council Policy Statement* (2014) and other applicable laws and regulations, has examined and approved the ethics application for the above-named research project.

Ethics approval is valid for the period indicated above and is subject to the conditions listed in the section entitled "Special Conditions or Comments". The "Renewal/Project Closure" form must be completedfour weeks before the above-referenced expiry date to request a renewal of this ethics approval or closure of the file.

Any changes made to the project must be approved by the REB before being implemented, except when necessary to remove participants from immediate endangerment or when the modification(s) only pertain to administrative or logistical components of the project. Investigators must also promptly alert the REB of any changes that increase the risk to participant(s), any changes that considerably affect the conduct of the project, all unanticipated and harmful events that occur, and new information that may negatively affect the conduct of the project or the safety of the participant(s).

Kim THOMPSON
Responsable d'éthique en recherche / Protocol Officer
Pour/For Barbara GRAVES Président(e) du/ Chair of the Comité d'éthique de la recherche en sciences sociales et humanités / Social Sciences and Humanities Research Ethics Board

550, rue Cumberland, pièce 154 Ottawa (Ontario) K1N 6N5 Canada
550 Cumberland Street, Room 154 Ottawa, Ontario K1N 6N5 Canada

613-562-5387 • 613-562-5338 • ethique@uOttawa.ca / ethics@uOttawa.ca
www.recherche.uottawa.ca/deontologie | www.recherche.uottawa.ca/ethics

## Appendix B: AEA Student E-Membership Approval to Access TIG Program Evaluators

**From:** info@eval.org <info@eval.org>
**Sent:** Sunday, October 10, 2021 7:45 PM
**To:**
**Subject:**

Attention : courriel externe | external email

### Order/Invoice Confirmation

American Evaluation Association

**Bill To:**

**Order Info:**

| Description | Qty | Price | Amount |
|---|---|---|---|
| Student E-Membership<br>1 Year<br>10/01/2021 - 09/30/2022 | 1 | $58.00 | $58.00 |
| | | Subtotal: | $58.00 |
| | | Tax: | $0.00 |
| | | Shipping: | $0.00 |
| | | Total Payment: | $58.00 |
| | | Credits: | $0.00 |
| | | Balance Due: | $0.00 |

## Appendix C: Phase 1 Program Director Survey – English

### Evaluation in Competency-Based Medical Education: A Mixed Methods Study

**1. I consent to completing this survey.**

O I agree. O I do not agree.

**2. Does your program engage in program evaluation?**

O Yes O No

If yes, go to question #3.
If no, go to question #9.

**N (Never); R (Rarely); S (Sometimes); F (Frequently); A (Always); DK (Don't Know)**

***Please select ONE response for each question.***

| | **3. Thinking about your program, how often do you or other individuals…** | **N** | **R** | **S** | **F** | **A** | **DK** |
|---|---|---|---|---|---|---|---|
| a | … work with a program evaluator? | □ | □ | □ | □ | □ | □ |
| b | … invite stakeholders (i.e., residents, faculty, physicians, community members) to participate in program evaluation(s)? | □ | □ | □ | □ | □ | □ |
| c | … clarify the purpose(s) of program evaluations? | □ | □ | □ | □ | □ | □ |
| d | … develop program evaluation questions? | □ | □ | □ | □ | □ | □ |
| e | … clarify program logic (i.e., how a program intends to work)? | □ | □ | □ | □ | □ | □ |
| f | … design a program evaluation proposal(s)/plan(s)? | □ | □ | □ | □ | □ | □ |
| g | … design data collection tools for program evaluation (e.g., surveys, interview guides)? | □ | □ | □ | □ | □ | □ |
| h | … collect quantitative program evaluation data? | □ | □ | □ | □ | □ | □ |
| i | … analyze quantitative program evaluation data? | □ | □ | □ | □ | □ | □ |
| j | … interpret quantitative program evaluation data? | □ | □ | □ | □ | □ | □ |
| k | … collect qualitative program evaluation data? | □ | □ | □ | □ | □ | □ |
| l | … analyze qualitative program evaluation data? | □ | □ | □ | □ | □ | □ |
| m | … interpret qualitative program evaluation data? | □ | □ | □ | □ | □ | □ |
| n | … use program evaluation findings? | □ | □ | □ | □ | □ | □ |
| o | Please share any other evaluation activities completed by you or other individuals of your program. | | | | | | |

| **N (Never); R (Rarely); S (Sometimes); F (Frequently); A (Always); DK (Don't Know)**<br>***Please select ONE response for each question.*** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **4. How often do you or other individuals…** | | **N** | **R** | **S** | **F** | **A** | **DK** |
| a | … evaluate the needs of your program? | □ | □ | □ | □ | □ | □ |
| b | … evaluate if the program addresses its learners' needs? | □ | □ | □ | □ | □ | □ |
| c | … evaluate if the program addresses its faculty's needs? | □ | □ | □ | □ | □ | □ |
| d | … evaluate how well your program is designed? | □ | □ | □ | □ | □ | □ |
| e | … evaluate how well your program is operating? | □ | □ | □ | □ | □ | □ |
| f | … evaluate if your program is implemented as intended? | □ | □ | □ | □ | □ | □ |
| g | … evaluate if learners are satisfied with the program? | □ | □ | □ | □ | □ | □ |
| h | … evaluate if faculty are satisfied with the program? | □ | □ | □ | □ | □ | □ |
| i | … evaluate your program's outcomes and impact? | □ | □ | □ | □ | □ | □ |
| j | … evaluate the program's cost and cost-effectiveness? | □ | □ | □ | □ | □ | □ |
| k | … investigate if program resources are used efficiently? | □ | □ | □ | □ | □ | □ |
| l | Please share any other program areas that you or other individuals evaluate. | | | | | | |

| **N (Never); R (Rarely); S (Sometimes); F (Frequently); A (Always); DK (Don't Know)**<br>***Please select ONE response for each question.*** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **5. How often do you or other individuals engage in program evaluation to...** | | **N** | **R** | **S** | **F** | **A** | **DK** |
| a | … satisfy accountability demands? | □ | □ | □ | □ | □ | □ |
| b | … produce evidence of program effectiveness? | □ | □ | □ | □ | □ | □ |
| c | … make decisions about the program? | □ | □ | □ | □ | □ | □ |
| d | … develop evaluation and inquiry skills? | □ | □ | □ | □ | □ | □ |
| e | … understand programming issues more fully? | □ | □ | □ | □ | □ | □ |
| f | … stimulate changes in clinical practice? | □ | □ | □ | □ | □ | □ |
| g | … stimulate changes in educational practice? | □ | □ | □ | □ | □ | □ |
| h | … learn about other people's perspectives of programs? | □ | □ | □ | □ | □ | □ |
| i | … improve its overall practices? | □ | □ | □ | □ | □ | □ |
| j | Please share any other reasons why you or other individuals are engaging or not engaging in program evaluation. | | | | | | |

| **SD (Strongly Disagree); D (Disagree); A (Agree); SA (Strongly Agree); DK (Don't Know)** *Select ONE response for each statement.* | | | | | | |
|---|---|---|---|---|---|---|
| **6.** | **Please indicate the extent to which you agree or disagree with each of the following statements.** | **SD** | **D** | **A** | **SA** | **DK** |
| a | My program uses program evaluation to better understand its functioning and practices. | □ | □ | □ | □ | □ |
| b | My program uses program evaluation to learn from its experiences. | □ | □ | □ | □ | □ |
| c | My program bases decisions about functioning on program evaluation results. | □ | □ | □ | □ | □ |
| d | My program uses program evaluation results to influence changes in program policies and procedures. | □ | □ | □ | □ | □ |
| e | My program uses program evaluation to comply with reporting demands. | □ | □ | □ | □ | □ |
| f | My program uses program evaluation to justify decisions previously made about its functioning. | □ | □ | □ | □ | □ |
| g | My program uses program evaluation to develop its research and inquiry skills. | □ | □ | □ | □ | □ |
| h | My program's behaviour and thinking changes as a result of engaging in program evaluation. | □ | □ | □ | □ | □ |
| i | Please share any other insight into why you or other individuals use or do not use program evaluation. | | | | | |

**7. Does your program have an employee whose primary responsibility is program evaluation?**

O Yes O No O I don't know.

**8. Does your program receive funding for program evaluation?**

O Yes O No O I don't know. O I prefer not to say.

**9. Why does your program not engage in program evaluation (please select all that apply)?**

- My program has no personnel to do program evaluation.
- My program has no funding to do program evaluation.
- My program has no time to do program evaluation.
- My program does not know how to do program evaluation.
- Other, please specify:

**10. Would you be interested in receiving a summary of the survey results?**

O Yes O No

*If a respondent selects "Yes" as their answer, they will be asked to provide their email address.*

**11. Would you be interested in participating in a follow-up interview for this study? You will be compensated with a $10 Amazon gift card for your participation in the follow-up interview.**

O Yes, please email additional information to me.

O Maybe, please email additional information to me.

O No, not at this time.

*If a respondent selects "Yes" or "Maybe" as their answer, they will be asked to provide their email address.*

**Thank you for your participation!**

## Appendix D: Phase 1 Program Director Survey – French

### L'évaluation à formation médicale fondée sur les compétences : une étude de méthodes mixtes

1. **Je consens à répondre à ce sondage.**
   O Oui, j'accepte de répondre aux questions de ce sondage.
   O Non, je n'accepte pas de répondre aux questions de ce sondage.

2. **Votre programme participe-t-il à l'évaluation de programmes?**
   O Oui O Non

Si vous avez répondu Oui, continuez à la question #3.
Si vous avez répondu Non, continuez à la question #9.

**J (Jamais); R (Rarement); P (Parfois); F (Fréquemment); T (Toujours); NSP (Ne sais pas)**
***Veuillez sélectionner UNE réponse pour chaque question.***

| | **3. En pensant à votre programme, à quelle fréquence est-ce que vous ou autres personnes…** | **J** | **R** | **P** | **F** | **T** | **NSP** |
|---|---|---|---|---|---|---|---|
| a | … travaillez/travaillent avec un évaluateur/une évaluatrice de programmes? | □ | □ | □ | □ | □ | □ |
| b | … invitez/invitent des parties prenantes (c.-à-d. des résidents, des professeurs, des médecins, des membres de la communauté) à participer dans l'évaluation (aux évaluations) de programmes? | □ | □ | □ | □ | □ | □ |
| c | … clarifiez/clarifient le but de l'évaluation de programmes? | □ | □ | □ | □ | □ | □ |
| d | … développez/développent des questions d'évaluation de programmes? | □ | □ | □ | □ | □ | □ |
| e | … clarifiez/clarifient la logique du programme (c.-à-d. comment un programme a l'intention de fonctionner)? | □ | □ | □ | □ | □ | □ |
| f | … concevez/conçoivent une(des) proposition(s) / un(des) plan(s) d'évaluation de programmes? | □ | □ | □ | □ | □ | □ |
| g | … concevez/conçoivent des outils de collecte de données pour l'évaluation de programmes (p. ex. des sondages, des guides d'entrevue)? | □ | □ | □ | □ | □ | □ |
| h | … recueillez/recueillent des données quantitatives pour l'évaluation de programmes? | □ | □ | □ | □ | □ | □ |
| i | … analysez/analysent des données quantitatives pour l'évaluation de programmes? | □ | □ | □ | □ | □ | □ |
| j | … interprétez/interprètent des données quantitatives pour l'évaluation de programmes? | □ | □ | □ | □ | □ | □ |
| k | … recueillez/recueillent des données qualitatives pour l'évaluation de programmes? | □ | □ | □ | □ | □ | □ |
| l | … analysez/analysent des données qualitatives pour l'évaluation de programmes? | □ | □ | □ | □ | □ | □ |
| m | … interprétez/interprètent des données qualitatives pour l'évaluation de programmes? | □ | □ | □ | □ | □ | □ |
| n | … utilisez/utilisent les résultats de l'évaluation de programmes? | □ | □ | □ | □ | □ | □ |
| o | Veuillez partager toute autre activité d'évaluation réalisée par vous ou autres personnes. | | | | | | |

| **J (Jamais); R (Rarement); P (Parfois); F (Fréquemment); T (Toujours); NSP (Ne sais pas)**<br>***Veuillez sélectionner UNE réponse pour chaque question.*** | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **4. À quelle fréquence est-ce que vous ou autres personnes…** | **J** | **R** | **P** | **F** | **T** | **NSP** |
| a | … évaluez/évaluent les besoins de votre/leur programme? | □ | □ | □ | □ | □ | □ |
| b | … évaluez/évaluent si le programme répond aux besoins de ses étudiant(e)s? | □ | □ | □ | □ | □ | □ |
| c | … évaluez/évaluent si le programme répond aux besoins de sa faculté? | □ | □ | □ | □ | □ | □ |
| d | … évaluez/évaluent la manière dont il est conçu? | □ | □ | □ | □ | □ | □ |
| e | … évaluez/évaluent son fonctionnement? | □ | □ | □ | □ | □ | □ |
| f | … évaluez/évaluent sa mise en oeuvre? | □ | □ | □ | □ | □ | □ |
| g | … évaluez/évaluent si les apprenants en sont satisfaits? | □ | □ | □ | □ | □ | □ |
| h | …évaluez/évaluent si la faculté en est satisfaite? | □ | □ | □ | □ | □ | □ |
| i | … évaluez/évaluent ses résultats et son effet? | □ | □ | □ | □ | □ | □ |
| j | … évaluez/évaluent son coût et son efficacité? | □ | □ | □ | □ | □ | □ |
| k | … enquêtez/enquêtent si les ressources sont utilisées efficacement? | □ | □ | □ | □ | □ | □ |
| l | Veuillez partager tous les autres domaines de votre programme évalués par vous ou autres personnes. | | | | | | |

| **J (Jamais); R (Rarement); P (Parfois); F (Fréquemment); T (Toujours); NSP (Ne sais pas)**<br>***Veuillez sélectionner UNE réponse pour chaque question.*** | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **5. À quelle fréquence est-ce que vous vous engagez ou autres personnes s'engagent dans l'évaluation de programme pour...** | **J** | **R** | **P** | **F** | **T** | **NSP** |
| a | … satisfaire des demandes de responsabilité? | □ | □ | □ | □ | □ | □ |
| b | … produire des preuves de l'efficacité du programme? | □ | □ | □ | □ | □ | □ |
| c | … faire des décisions concernant le programme? | □ | □ | □ | □ | □ | □ |
| d | … développer des compétences d'évaluation et d'enquête? | □ | □ | □ | □ | □ | □ |
| e | … comprendre plus complètement des problèmes de programmation? | □ | □ | □ | □ | □ | □ |
| f | … stimuler des changements dans la pratique clinique? | □ | □ | □ | □ | □ | □ |
| g | … stimuler des changements dans la pratique pédagogique? | □ | □ | □ | □ | □ | □ |
| h | … apprendre le point de vue d'autres personnes du programme? | □ | □ | □ | □ | □ | □ |
| i | … améliorer des pratiques générales? | □ | □ | □ | □ | □ | □ |
| j | Veuillez indiquer toute autre raison pour laquelle vous engagez ou autres personnes s'engagent dans l'évaluation de programmes. | | | | | | |

| **FD (Fortement en désaccord); D (Désaccord); A (D'accord); FA (Fortement d'accord); NSP (Ne sais pas)** *Veuillez sélectionner UNE réponse pour chaque question.* | | | | | | |
|---|---|---|---|---|---|---|
| | **6. Veuillez indiquer dans quelle mesure vous êtes en accord ou en désaccord avec chacune des affirmations suivantes.** | **FD** | **D** | **A** | **FA** | **NSP** |
| a | Mon programme utilise l'évaluation de programmes pour mieux comprendre le programme et ses pratiques. | □ | □ | □ | □ | □ |
| b | Mon programme utilise l'évaluation de programmes pour apprendre des leçons de ses expériences. | □ | □ | □ | □ | □ |
| c | Mon programme fonde ses décisions sur les résultats de l'évaluation de programmes. | □ | □ | □ | □ | □ |
| d | Mon programme utilise l'évaluation de programmes pour informer les politiques et les procédures du programme. | □ | □ | □ | □ | □ |
| e | Mon programme utilise l'évaluation de programmes pour se conformer aux exigences de rapports. | □ | □ | □ | □ | □ |
| f | Mon programme utilise l'évaluation de programmes pour justifier les décisions précédemment prises par le programme. | □ | □ | □ | □ | □ |
| g | Mon programme utilise l'évaluation de programmes pour développer ses capacités de recherche et d'enquête. | □ | □ | □ | □ | □ |
| h | Le comportement et la logique de mon programme changent en fonction de l'évaluation de programmes. | □ | □ | □ | □ | □ |
| i | Veuillez partager toute autre information expliquant pourquoi vous utilisez ou n'utilisez pas ou autres personnes utilisent ou n'utilisent pas l'évaluation de programmes. | | | | | |

**7. Votre programme a-t-il un employé dont la responsabilité principale est l'évaluation de programmes?**

O Oui O Non O Je ne sais pas.

**8. Votre programme reçoit-il des fonds pour l'évaluation de programmes?**

O Oui O Non O Je ne sais pas O Je préfère ne pas le dire.

**9. Pourquoi est-ce que votre programme ne s'engage pas dans l'évaluation de programmes (sélectionnez tout ce qui s'applique)?**

- Mon programme n'y a personne pour faire l'évaluation de programmes.
- Mon programme n'a pas de financement pour faire l'évaluation de programmes.
- Mon programme n'a pas de temps pour faire l'évaluation de programmes.
- Mon programme ne sait pas comment faire l'évaluation de programmes.
- Autre, veuillez préciser:

**10. Seriez-vous intéressé à recevoir un résumé des résultats de sondage?**

O Oui O Non

*Si un répondant choisit «Oui» comme réponse, il sera demandé de fournir son courriel.*

**11. Seriez-vous intéressé à participer à une entrevue de suivi pour cette étude? Vous serez compensé par une carte-cadeau d'Amazon (10 $) pour votre participation à l'entrevue de suivi.**

O Oui, veuillez m'envoyer des informations supplémentaires par courriel.

O Peut-être, veuillez m'envoyer des informations supplémentaires par courriel.

O Non, pas pour le moment.

*Si un répondant choisit «Oui» ou «Peut-être» comme réponse, il sera demandé de fournir son adresse électronique.*

**Merci pour votre participation!**

## Appendix E: Phase 1 Information Letter – English

uOttawa

Université d'Ottawa
Faculté d'éducation

University of Ottawa
Faculty of Education

145, Jean-Jacques Lussier
Ottawa ON K1N 6N5
Canada

www.education.uOttawa.ca

Title of the study: Program Evaluation in Competency-Based Medical Education: A Mixed Methods Study



| **Student Investigator:** | **Supervisor:** |
|---|---|
| Jenna Milosek (PhD candidate)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario | Katherine A. Moreau, PhD<br>Associate Professor, Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario |



Dear Program Director,

You are invited to take part in Phase 1 of this three-phase research study conducted by Jenna Milosek, PhD candidate from the University of Ottawa, under the supervision of Professor Katherine Moreau. You have been identified as a potential participant in this study because you are a Program Director for a Royal College of Physicians and Surgeons of Canada medical education program.

The purpose of this study is to explore the extent to which Competence by Design (CBD) programs are engaging in program evaluation, the ways in which these programs are engaging in program evaluation, and the reasons why they are engaging or not engaging in program evaluation. It also explores the barriers and facilitators for doing program evaluation as well as ways that CBD programs can build and support their capacities to do program evaluation and use program evaluation findings.

For Phase 1 of this study, you are invited to complete a short survey. The survey includes questions about the extent that your CBD program is engaging in program evaluation and the reasons why it is engaging or is not engaging in program evaluation.

The survey will take approximately **10 minutes** to complete. We ask that you complete the survey online using the following link [https://www.surveymonkey.ca/r/TLNLDKD] by **June 14, 2021**. Participants are asked to provide consent by clicking on "I agree" prior to accessing the survey.

You do not have to respond to any items or answer questions that make you feel uncomfortable. You may or may not benefit directly from the study. However, the findings from it will provide you with information on the extent and ways that CBD programs are engaging in program evaluation. It will also help you understand why programs are engaging or not engaging in program evaluation as well as the barriers and facilitators of doing program evaluation. Moreover, it will provide you with strategies for building and supporting your program's capacities to do program evaluation and use evaluation findings. To thank you for your participation in the study, you can receive a summary of the findings.

The survey is designed so that your identity will remain strictly confidential. Your responses to the survey will only be used for understanding the extent to which CBD programs are engaging in program evaluation and the reasons why they are engaging or are not engaging in program evaluation.

Your survey responses will be combined with other responses so that you cannot be identified in published reports or presentations. Also, any written responses that could potentially reveal your identity (e.g., name, university, hospital, or region) will be removed from the survey and not included in the database.

Completed surveys will be stored on a password-protected computer. Only the members of the above-mentioned research team will have access to the surveys. Data will be conserved for five years after the publication of research findings. After this time, data will be deleted, shredded, and appropriately discarded.

The completion of this survey is **voluntary**. You can refuse to answer any questions without experiencing any negative consequences. In the future, if you are interested, you will receive an email inviting you to participate in Phase 2 of this study. Your participation in this second phase is **optional.** Phase 2 involves a one-on-one interview with Jenna Milosek. The interview will focus on how CBD programs are engaging in program evaluation, the barriers and facilitators of engaging in program evaluation, and how CBD programs can build their capacities to do program evaluation and use evaluation findings. The interview will last approximately 1 hour and will be scheduled at a time that is convenient for you. Additional information will be sent to you by email, if you are interested, inviting you to participate in Phase 2 of this research study. **By completing this survey for Phase 1, you are <u>not</u> consenting to participating in Phase 2.**

If you have any questions about the study, please contact Jenna Milosek or Katherine Moreau at the coordinates below. If you have any questions regarding the ethical conduct of this study, you may contact the Ethics Office at (613-562-5387, room 154).

Sincerely,

| | |
|---|---|
| _____ | _____ |
| Jenna Milosek (PhD Candidate)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario | Katherine Moreau, PhD (Supervisor)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario |

## Appendix F: Phase 1 Information Letter – French

uOttawa

Université d'Ottawa
Faculté d'éducation

University of Ottawa
Faculty of Education

145, Jean-Jacques Lussier
Ottawa ON K1N 6N5
Canada

www.education.uOttawa.ca

Titre d'étude: L'évaluation de programmes dans la formation médicale fondée sur les compétences : une étude à méthodes mixtes

| **Enquêteuse:** | **Superviseuse:** |
|---|---|
| Jenna Milosek (candidate au doctorat)<br>Faculté d'éducation<br>Université d'Ottawa<br>Ottawa, Ontario | Katherine A. Moreau, Ph. D.<br>Professeure agrégée, Faculté d'éducation<br>Université d'Ottawa<br>Ottawa, Ontario |

Cher directeur/Chère directrice de programme,

Vous êtes invité à participer à la 1ère phase de cette étude de recherche à trois phases menée par Jenna Milosek, doctorante de l'Université d'Ottawa, sous la direction de la professeure Katherine Moreau. Vous avez été identifié comme participant potentiel à cette étude parce que vous êtes un directeur/une directrice de programme pour un programme d'éducation médicale du Collège royal des médecins et chirurgiens du Canada.

Le but de cette étude est d'explorer la mesure dans laquelle les programmes de la compétence par conception (CPC) s'engagent dans l'évaluation de programmes, les façons dont ces programmes s'engagent dans l'évaluation de programmes, et les raisons pour lesquelles ils s'engagent ou ne s'engagent pas dans l'évaluation de programmes. Elle explore également les obstacles et les facilitateurs pour faire l'évaluation de programmes ainsi que les moyens par lesquels les programmes de CPC peuvent renforcer et utiliser les résultats d'une évaluation de programme.

Pour la 1ère phase de cette étude, vous êtes invité à répondre à un court sondage. Le sondage comprend des questions sur la mesure dans laquelle votre programme de CPC participe dans l'évaluation de programmes et les raisons pour lesquelles il s'engage ou ne s'engage pas dans l'évaluation de programmes.

Le sondage vous prendra environ **10 minutes** à compléter. Nous vous demandons de remplir le sondage en ligne en utilisant le lien suivant [https://fr.surveymonkey.ca/r/TLNQ7SK] avant le **14 juin 2021**. Les participants seront invités à donner leur consentement en cliquant sur « j'accepte » avant d'accéder au sondage.

Vous n'êtes pas obligé de répondre à des éléments ou à répondre aux questions qui vous mettent mal à l'aise. Vous pourriez bénéficier ou vous ne pourriez pas bénéficier directement de l'étude. Cependant, ses conclusions vous fourniront des informations sur l'étendue et la manière dont les programmes de CPC s'engagent dans l'évaluation de programmes. Cela vous aidera également à comprendre pourquoi les programmes s'engagent ou ne s'engagent pas dans l'évaluation de programmes ainsi que les obstacles et les facilitateurs à faire l'évaluation de programmes. De plus, il vous fournira des stratégies pour renforcer et utiliser les résultats d'une évaluation de programme. Pour vous remercier de votre participation à l'étude, vous pouvez recevoir un résumé des résultats.

Le sondage est conçu pour que votre identité reste strictement confidentielle. Vos réponses au sondage ne seront utilisées que pour comprendre dans quelle mesure les programmes de CPC s'engagent dans l'évaluation de programmes et les raisons pour lesquelles ils s'engagent ou ne s'engagent pas dans l'évaluation de programmes.

Vos réponses de sondage seront combinées avec d'autres réponses afin que vous ne puissiez pas être identifié aux rapports publiés ou aux présentations. De plus, toutes les réponses écrites susceptibles de révéler votre identité (par exemple, le nom, l'université, l'hôpital ou la région) seront supprimées du sondage et ne seront pas incluses dans la base de données.

Les sondages complétés seront gardés sur un ordinateur protégé par mot de passe. Seuls les membres de l'équipe de recherche susmentionnée auront accès aux sondages. Les données seront conservées pendant cinq ans après la publication des résultats de recherche. Après ce temps, les données seront supprimées, déchiquetées et éliminées convenablement.

L'achèvement de ce sondage est **volontaire**. Vous pouvez refuser de répondre à toutes les questions sans subir de conséquences négatives. À l'avenir, si vous êtes intéressé, vous recevrez un courriel vous invitant à participer à la 2e phase de cette étude. Votre participation à cette deuxième phase est **facultative**. La 2e phase comprend une entrevue individuelle avec Jenna Milosek. L'entrevue s'intéresse à la manière dont les programmes de CPC s'engagent dans l'évaluation de programmes, des obstacles et des facilitateurs de la participation dans l'évaluation de programmes, et comment des programmes de CPC peuvent renforcer leurs capacités à faire l'évaluation de programmes et à utiliser les résultats d'une évaluation de programme. L'entrevue durera environ 1 heure et sera cédulée à une heure qui vous convient. Des informations supplémentaires seront envoyées par courriel, si vous êtes intéressé, vous invitant à participer à la 2e phase de cette étude de recherche. **En répondant à ce sondage pour la 1ère phase, vous <u>n'acceptez pas</u> de participer à la 2e phase.**

Si vous avez des questions à propos de l'étude, veuillez contacter Jenna Milosek ou Katherine Moreau aux coordonnées ci-dessous. Si vous avez des questions concernant la conduite éthique de cette étude, vous pouvez contacter le Bureau d'éthique au (613-562-5387, bureau 154).

Sincèrement,


| | |
|---|---|
| Jenna Milosek (candidate au doctorat)<br>Faculté d'éducation<br>Université d'Ottawa<br>Ottawa, Ontario | Katherine A. Moreau, Ph. D. (superviseuse)<br>Professeure agrégée, Faculté d'éducation<br>Université d'Ottawa<br>Ottawa, Ontario |

**Appendix G: Phase 1 Number of Survey Respondents**

| | Survey question subject area | Number of respondents for a given question | Number of respondents who answered question | Number of missing respondents |
|---|---|---|---|---|
| Q1. | Survey respondents' consent to completing this survey. | 152 | 152 | 0 |
| Q2. | The consenting survey respondents whose program engages in program evaluation. | 149 | 143 | 6 |
| Q3a. | Survey respondents work with a program evaluator. | 127 | 86 | 41 |
| Q3b. | Survey respondents invite stakeholders to participate in program evaluations. | 127 | 90 | 37 |
| Q3c. | Survey respondents clarify the purposes of program evaluations. | 127 | 89 | 38 |
| Q3d. | Survey respondents develop program evaluation questions. | 127 | 89 | 38 |
| Q3e. | Survey respondents clarify program logic how a program intends to work. | 127 | 89 | 38 |
| Q3f. | Survey respondents design a program evaluation proposal(s)/plan(s). | 127 | 89 | 38 |
| Q3g. | Survey respondents design data collection tools for program evaluation. | 127 | 89 | 38 |
| Q3h. | Survey respondents collect quantitative program evaluation data. | 127 | 89 | 38 |
| Q3i. | Survey respondents analyze quantitative program evaluation data. | 127 | 89 | 38 |
| Q3j. | Survey respondents interpret quantitative program evaluation data. | 127 | 88 | 39 |
| Q3k. | Survey respondents collect qualitative program evaluation data. | 127 | 87 | 40 |
| Q3l. | Survey respondents analyze qualitative program evaluation data. | 127 | 88 | 39 |
| Q3m. | Survey respondents interpret qualitative program evaluation data. | 127 | 87 | 40 |
| Q3n. | Survey respondents use program evaluation findings. | 127 | 87 | 40 |
| Q3o. | Survey respondents share any other evaluation activities completed by them or other individuals of their program. | 127 | 17 | 110 |
| Q4a. | Survey respondents evaluate the needs of their program. | 127 | 86 | 41 |
| Q4b. | Survey respondents evaluate if the program addresses its learners' needs. | 127 | 86 | 41 |

| | Survey question subject area | Number of respondents for a given question | Number of respondents who answered question | Number of missing respondents |
|---|---|---|---|---|
| Q4c. | Survey respondents evaluate if the program addresses its faculty's needs. | 127 | 86 | 41 |
| Q4d. | Survey respondents evaluate how well their program is designed. | 127 | 85 | 42 |
| Q4e. | Survey respondents evaluate how well their program is operating. | 127 | 86 | 41 |
| Q4f. | Survey respondents evaluate if their program is implemented as intended. | 127 | 86 | 41 |
| Q4g. | Survey respondents evaluate if learners are satisfied with the program. | 127 | 85 | 42 |
| Q4h. | Survey respondents evaluate if faculty are satisfied with the program. | 127 | 86 | 41 |
| Q4i. | Survey respondents evaluate their program's outcomes and impact. | 127 | 86 | 41 |
| Q4j. | Survey respondents evaluate their program's cost and cost-effectiveness. | 127 | 86 | 41 |
| Q4k. | Survey respondents investigate if program resources are used efficiently. | 127 | 86 | 41 |
| Q4l. | Survey respondents share any other program areas that they or other individuals evaluate. | 127 | 7 | 120 |
| Q5a. | Survey respondents engage in program evaluation to satisfy accountability demands. | 127 | 81 | 46 |
| Q5b. | Survey respondents engage in program evaluation to produce evidence of program effectiveness. | 127 | 81 | 46 |
| Q5c. | Survey respondents engage in program evaluation to make decisions about the program. | 127 | 82 | 45 |
| Q5d. | Survey respondents engage in program evaluation to develop evaluation and inquiry skills. | 127 | 82 | 45 |
| Q5e. | Survey respondents engage in program evaluation to understand programming issues more fully. | 127 | 82 | 45 |
| Q5f. | Survey respondents engage in program evaluation to stimulate changes in clinical practice. | 127 | 82 | 45 |
| Q5g. | Survey respondents engage in program evaluation to stimulate changes in educational practice. | 127 | 82 | 45 |
| Q5h. | Survey respondents engage in program evaluation to learn about other people’s perspectives of programs. | 127 | 82 | 45 |
| Q5i. | Survey respondents engage in program evaluation to improve its overall practices. | 127 | 82 | 45 |

| | Survey question subject area | Number of respondents for a given question | Number of respondents who answered question | Number of missing respondents |
|---|---|---|---|---|
| Q5j. | Survey respondents share any other reasons why they or other individuals are engaging or not engaging in program evaluation. | 127 | 6 | 121 |
| Q6a. | Survey respondents' program uses program evaluation to better understand its functioning and practices. | 127 | 80 | 47 |
| Q6b. | Survey respondents' program uses program evaluation to learn from its experiences. | 127 | 80 | 47 |
| Q6c. | Survey respondents' program bases decisions about functioning on program evaluation results. | 127 | 80 | 47 |
| Q6d. | Survey respondents' program uses program evaluation results to influence changes in program policies and procedures. | 127 | 80 | 47 |
| Q6e. | Survey respondents' program uses program evaluation to comply with reporting demands. | 127 | 80 | 47 |
| Q6f. | Survey respondents' program uses program evaluation to justify decisions previously made about its functioning. | 127 | 79 | 48 |
| Q6g. | Survey respondents' program uses program evaluation to develop its research and inquiry skills. | 127 | 80 | 47 |
| Q6h. | The behaviour and thinking of survey respondents' program changes as a result of engaging in program evaluation. | 127 | 80 | 47 |
| Q6i. | Survey respondents share any other insight into why they or other individuals use or do not use program evaluation. | 127 | 4 | 123 |
| Q7. | Survey respondents whose program has an employee whose primary responsibility is program evaluation. | 127 | 80 | 47 |
| Q8. | Survey respondents whose program receives funding for program evaluation. | 127 | 81 | 46 |
| Q9a. | Survey respondents whose program does not engage in program evaluation because the program has no personnel to do program evaluation. | 16 | 13 | 3 |
| Q9b. | Survey respondents whose program does not engage in program evaluation because the program has no funding to do program evaluation. | 16 | 13 | 3 |

| | Survey question subject area | Number of respondents for a given question | Number of respondents who answered question | Number of missing respondents |
|---|---|---|---|---|
| Q9c. | Survey respondents whose program does not engage in program evaluation because the program has no time to do program evaluation. | 16 | 13 | 3 |
| Q9d. | Survey respondents whose program does not engage in program evaluation because the program does not know how to do program evaluation. | 16 | 13 | 3 |
| Q9e. | Survey respondents specify any other reasons why their program does not engage in program evaluation. | 16 | 6 | 10 |
| Q10a. | Survey respondents interested or not interested in receiving a summary of the survey results. | 149 | 93 | 56 |
| Q10b. | Survey respondents interested in receiving a summary of the survey results who provided his or her email address. | 60 | 57 | 3 |
| Q11a. | Survey respondents interested (“yes” or “maybe”) or not interested in participating in a follow-up interview. | 149 | 92 | 57 |
| Q11b. | Survey respondents interested (“yes” or “maybe”) in participating in a follow-up interview who provided his or her email address. | 92 | 36 | 56 |

## Appendix H: Phase 2 Interview Guides

### Program Directors Who Engage in Program Evaluation Interview Guide

***For use with program directors who indicated that they DO use program evaluation to evaluate their CBD program***.

DATE: _____

TIME: _____

PARTICIPANT ID: _____

INTERVIEWER: _____

We want to understand the extent to which Competence by Design (CBD) Canadian residency programs are engaging in program evaluation, the ways in which these programs are engaging in program evaluation, and the reasons why they are engaging or not engaging in program evaluation. We also want to explore the barriers and facilitators for doing program evaluation as well as ways that CBD programs can build and support their capacities to do program evaluation and use program evaluation findings.

In this interview, we often refer to **program evaluation**. This is defined as the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming.

To begin, I would like to confirm some information to ensure accurate context for the interview.

**You are currently the Program Director for the following specialty: _____**

**Your program has been implementing CBD since its launch date in: _____**

On your survey, you indicated that your program **engages** in program evaluation.

| ***Research Question 3: How are CBD programs engaging in program evaluation?*** |
|---|
| **1.** How do you and/or your team evaluate your CBD program?<br><br>- Who is involved in the evaluation of your CBD program?<br>- How is the evaluation designed and implemented?<br>- How are data collected?<br>- What types of data are collected?<br>- How are data analyzed?<br>- How are evaluation findings disseminated/reported?<br>- How are evaluation findings used? |

| |
|---|
| **2.** What do you and/or your team hope to learn from the evaluation of your CBD program?<br><br>- What are the reasons why you and/or your team engage in program evaluation of your CBD program? |

| ***Research Question 4 (1st half): How can CBD programs build their capacities to do program evaluation?*** |
|---|
| **3.** How confident are you in terms of your abilities and/or your team's abilities to evaluate your CBD program? |
| **4.** What training have you and/or your team received in program evaluation? |
| **5.** What types of resources do you and/or your team have to help evaluate your CBD program? |
| **6.** How does access to training or the availability of resources (e.g., human, financial, time) affect your ability and/or your team's abilities to evaluate your CBD program evaluation? |
| **7.** What do you, your team, and/or your program need to build its capacity to do program evaluation? |
| **8.** If you could make changes in the way your CBD program is evaluated, what would these changes be? |

| ***Research Question 4 (2nd half): How can CBD programs build their capacities to use evaluation findings?*** |
|---|
| **9.** How do you and/or your team use program evaluation to better understand your program and its practices?<br><br>- How do you and/or your team use evaluation results to influence changes in program policies and procedures?<br><br>- How do you and/or your team use evaluation to comply with reporting demands? |
| **10.** How do you and/or your team ensure that your program stakeholders can use its program evaluation findings?<br><br>- What do you and/or your team need to ensure that your program stakeholders can use its program evaluation findings? |

**11.** Would you be interested in receiving a summary of the findings?
O Yes O No

*Or, in your survey response, you already responded 'YES' to receiving a summary of the findings.*
*Is this still your email address?* _____

*Or, in your survey response, you already responded 'NO' to receiving a summary of the findings. Are you sure, would you like to change your mind?*
*What is your email address?* _____

**12.** Thank you for participating in this follow-up interview. You will be compensated with a $10 Amazon gift card for your participation in the follow-up interview.

*Confirm email address:* _____

**13.** I will email you a transcript of this interview for you to review and make any changes in your responses. I will email you in the next day or two. Please review it and make any changes within two weeks.

**Thank you for your participation!**

## Program Directors Who do not Engage in Program Evaluation Interview Guide

***For use with program directors who indicated that they do <u>NOT</u> use program evaluation to evaluate their CBD program***.

DATE: _____

TIME: _____

PARTICIPANT ID: _____

INTERVIEWER: _____

We want to understand the extent to which Competence by Design (CBD) Canadian residency programs are engaging in program evaluation, the ways in which these programs are engaging in program evaluation, and the reasons why they are engaging or not engaging in program evaluation. We also want to explore the barriers and facilitators for doing program evaluation as well as ways that CBD programs can build and support their capacities to do program evaluation and use program evaluation findings.

In this interview, we often refer to **program evaluation**. This is defined as the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming.

To begin, I would like to confirm some information to ensure accurate context for the interview.

**You are currently the Program Director for the following specialty: _____**

**Your program has been implementing CBD since its launch date in: _____**

On your survey, you indicated that your program **<u>does not engage</u>** in program evaluation.

| ***Research Question 3: How are CBD programs engaging in program evaluation?*** |
|---|
| **1.** Why does your CBD program not engage in program evaluation?<br><br>- What are the barriers preventing your program from engaging in program evaluation? |
| **2.** Are you and/or your team interested in evaluating your CBD program?<br><br>- Why are you and/or your team interested in evaluating your CBD program?<br><br>- How often would you and/or your team like to engage in program evaluation of your CBD program? |

| |
|---|
| - Who would you and/or your team like to involve in the evaluation of your CBD program? |
| **3.** What program areas may benefit from being evaluated? |

| ***Research Question 4 (1st half): How can CBD programs build their capacities to do program evaluation?*** |
|---|
| **4.** How confident are you in terms of your abilities and/or your team's abilities to evaluate your CBD program? |
| **5.** What do you, your team, and/or your program need to build its capacity to do program evaluation? |
| **6.** What resources would you and/or your team need to be able to evaluate your CBD program? |
| **7.** How does access to training or the availability of resources (e.g., human, financial, time) affect your ability and/or your team's abilities to evaluate your CBD program evaluation? |

| ***Research Question 4 (2nd half): How can CBD programs build their capacities to use evaluation findings?*** |
|---|
| **8.** How would you and/or your team use program evaluation to better understand your program and its practices?<br><br>- How would you and/or your team use evaluation results to influence changes in program policies and procedures?<br><br>- How would you and/or your team use evaluation to comply with reporting demands? |
| **9.** What would you and/or your team need to ensure that your program stakeholders could use its program evaluation findings? |

**10.** Would you be interested in receiving a summary of the findings?
O Yes O No

*Or, in your survey response, you already responded 'YES' to receiving a summary of the findings.*
*Is this still your email address?* _____

*Or, in your survey response, you already responded 'NO' to receiving a summary of the findings.*
*Are you sure, would you like to change your mind?*
*What is your email address?* _____

**11.** Thank you for participating in this follow-up interview. You will be compensated with a $10 Amazon gift card for your participation in the follow-up interview.

*Confirm email address:* _____

**12.** I will email you a transcript of this interview for you to review and make any changes in your responses. I will email you in the next day or two. Please review it and make any changes within two weeks.

**Thank you for your participation!**

## Appendix I: Phase 2 Information Letter

uOttawa

Title of the study: Program Evaluation in Competency-Based Medical Education: A Mixed Methods Study


| Student Investigator: | Supervisor: |
|---|---|
| Jenna Milosek (PhD candidate)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario | Katherine A. Moreau, PhD<br>Associate Professor, Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario |

Université d'Ottawa
Faculté d'éducation

University of Ottawa
Faculty of Education


Dear Program Director,

You are invited to take part in Phase 2 of this three-phase research study by elaborating on your experiences with program evaluation within your medical education program. The purpose of this study is to explore the extent to which Competence by Design (CBD) programs are engaging in program evaluation, the ways in which these programs are engaging in program evaluation, and the reasons why they are engaging or not engaging in program evaluation. It also explores the barriers and facilitators for doing program evaluation as well as ways that CBD programs can build and support their capacities to do program evaluation and use program evaluation findings.

If you agree to participate in this phase of the study, you will be asked to take part in a one-on-one interview. The online interview will be conducted remotely. The interview will take place using Zoom at a time that is convenient for you. We will ask that you select a location for the interview that is private and convenient for you. Jenna Milosek will then conduct the interview with you from a private office. The interview will be a semi-structured conversation-style session. In the interview, you will be asked open-ended questions on how your CBD program is engaging in program evaluation, the barriers and facilitators of engaging in program evaluation, and how CBD programs can build their capacities to do program evaluation and use evaluation findings. The interview will take approximately **60 minutes** to complete. The interview will be audio-recorded and transcribed by Jenna Milosek for analysis. Participants will be selected on a first-come, first-served basis.

Some of the questions in the interview may make you feel uncomfortable because they ask about how your CBD program is engaging in program evaluation. You do not have to respond to any questions that make you feel uncomfortable. You may refuse to participate in the study. Your responses, recordings and transcripts will remain confidential.

You may or may not benefit directly from this study. However, the findings from it will provide you with information on the extent and ways that CBD programs are engaging in program evaluation. It will also help you understand why programs are engaging or not engaging in program evaluation as well as the barriers and facilitators of doing program evaluation. Moreover, it will provide you with strategies for building and supporting my program's capacities to do program evaluation and use evaluation findings. To thank you for your participation in the interview phase of the study, you will receive a summary of the findings and a $10 Amazon gift card upon submission of your email address.

145, Jean-Jacques Lussier
Ottawa ON K1N 6N5
Canada

www.education.uOttawa.ca

**If you are interested in participating in a one-on-one interview for this study, please contact Jenna Milosek by July 31, 2021 to obtain additional information about the interview.** As mentioned, the interview will be scheduled at a time that is convenient for you. Your written consent will be obtained prior to the interview. We will email you a copy of the consent form to review prior to the interview.

If you have any questions about the study, please contact Jenna Milosek or Katherine Moreau at the coordinates below. If you have any questions regarding the ethical conduct of this study, you may contact the Ethics Office at (613-562-5387, room 154).

Sincerely,

| | |
|---|---|
| _____ | _____ |
| Jenna Milosek (PhD Candidate)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario | Katherine Moreau, PhD (Supervisor)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario |

## Appendix J: Phase 2 Consent Form

uOttawa

Université d'Ottawa
Faculté d'éducation

University of Ottawa
Faculty of Education

Title of the study: Program Evaluation in Competency-Based Medical Education: A Mixed Methods Study



| **Student Investigator:** | **Supervisor:** |
|---|---|
| Jenna Milosek (PhD candidate)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario | Katherine A. Moreau, PhD<br>Associate Professor, Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario |



**Invitation to Participate:** I am invited to participate in a one-on-one interview for Phase 2 of this three-phase research study conducted by Jenna Milosek, PhD (Education) candidate from the University of Ottawa, under the supervision of Professor Katherine Moreau.

**Purpose of the Study:** The purpose of this study is to explore the extent to which Competence by Design (CBD) programs are engaging in program evaluation, the ways in which these programs are engaging in program evaluation, and the reasons why they are engaging or not engaging in program evaluation. It also explores the barriers and facilitators for doing program evaluation as well as ways that CBD programs can build and support their capacities to do program evaluation and use program evaluation findings.

**Participation:** My participation will consist of taking part in one one-on-one interview. The online interview will be conducted remotely. The interview will take place via Zoom at a time that is convenient for me. I will be asked to select a location for the interview that is private and convenient for me. Jenna Milosek will then conduct the interview with me from a private office. The interview will be a semi-structured conversation-style session. In the interview, I will be asked open-ended questions on how my CBD program is engaging in program evaluation, the barriers and facilitators of engaging in program evaluation, and how CBD programs can build their capacities to do program evaluation and use evaluation findings. The interview will take approximately **60 minutes** to complete. I understand that the interview will be audio-recorded and transcribed by Jenna Milosek for analysis. I understand that participants will be selected on a first come, first served basis.

**Benefits:** I may or may not benefit directly from this study. However, the findings from it will provide me with information on the extent and ways that CBD programs are engaging in program evaluation. It will also help me understand why programs are engaging or not engaging in program evaluation as well as the barriers and facilitators of doing program evaluation. Moreover, it will provide me with strategies for building and supporting my program's capacities to do program evaluation and use evaluation findings. To thank me for my participation in the interview phase of the study, I will receive a summary of the findings and a $10 Amazon gift card upon submission of my email address.

145, Jean-Jacques Lussier
Ottawa ON K1N 6N5
Canada

www.education.uOttawa.ca

**Confidentiality and anonymity:** I have received assurance from the researchers that the information I will share will remain strictly confidential. I understand that the information I provide in the interview will only be used for describing how CBD programs are engaging in program evaluation, the barriers and facilitators of engaging in program evaluation, and how CBD programs can build their capacities to do program evaluation and use evaluation findings. Only Jenna Milosek and her supervisor will know my identity, and I will not be asked to state my name in the interview. Any information that could potentially reveal my identity (e.g., name, university, hospital) will be eliminated from the audio recording and transcript so that I cannot be identified in published reports or presentations. Thus, anonymity can be guaranteed in the dissemination of the findings. Jenna Milosek will email me a copy of my interview transcript to review and revise prior to analyses. I will have two weeks to review my transcript and share any necessary changes. Jenna Milosek will ensure that the transcript is secured on her password-protected laptop.

**Conservation of data:** The digital audio recording of the interview will be downloaded and erased from the audio-recorder immediately after the interview. All data will be stored on a password-protected laptop. Only the members of the above-mentioned research team will have access to the data. Data will be conserved for five years after the publication of research findings. After this time, data will be deleted, shredded, and appropriately discarded.

**Voluntary Participation:** I am under no obligation to participate and if I choose to participate, I can withdraw from the study at any time and/or refuse to answer any questions without experiencing any negative consequences. If I choose to withdraw, all data gathered until the time of withdrawal will be deleted, destroyed, and not included in any publications or presentations.

**Acceptance:** I _____, agree to participate in the above-mentioned research study conducted by Jenna Milosek PhD (Education) candidate from the Faculty of Education, University of Ottawa under the supervision of Professor Katherine Moreau.

If I have any questions about the study, I may contact Jenna Milosek or Katherine Moreau at:

| | |
|---|---|
| _____ | _____ |
| Jenna Milosek (PhD candidate)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario | Katherine Moreau, PhD (Supervisor)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario |

If I have any questions regarding the ethical conduct of this study, I may contact the Ethics Office at (613-562-5387, room 154).

I will save or print a copy of this form for my personal records.

Participant's signature: _____ Date: _____

Researcher's signature: _____ Date: _____

## Appendix K: Phase 2 Final NVivo Codes

**Program directors of CBD programs are engaging in program evaluation by:**

- Struggling to engage in program evaluation because of:
  - Limited available resources
  - Buy-in
- Using ad hoc evaluation methods
- Using a team-based format

**Program directors of CBD programs can build their capacities to do program evaluation by:**

- Developing expertise in their program evaluation abilities
- Acquiring program evaluation resources for guidance
- Advocating for clear program evaluation expectations from:
  - The Royal College
  - The PGME office
  - Within their specialties

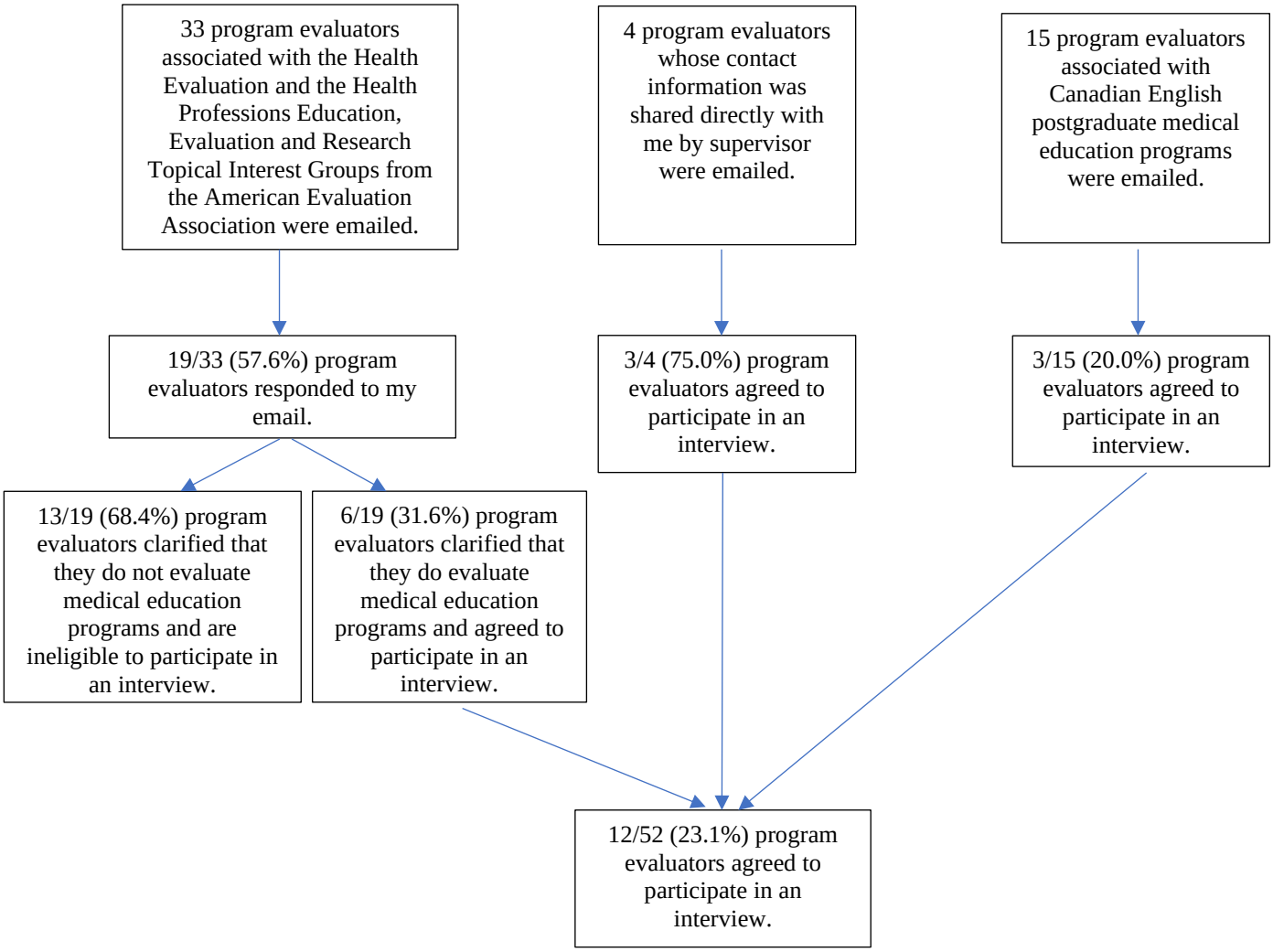
**Program directors of CBD programs can build their capacities to use evaluation findings by:**

- Asking meaningful questions that are appropriate to the stage of the program
- Establishing buy-in for program evaluation to increase the likeliness of using the findings
- Sharing findings in ways that encourage use with:
  - Faculty
  - Residents
  - Other programs
- Understand evaluation as helpful for the accreditation process

**Appendix L: Phase 2 Sample Data Analysis Matrix**

| Factor | Quotations |
|---|---|
| Struggling to engage in program evaluation because of: | |
| Limited available resources | "The simple answer [is] it's time and money. Unfortunately" (P1). |
| | "CBD came in, but no extra money came in. And, so, I think programs in general have found that CBD costs a lot more than what the Royal College was suggesting it would" (P3). |
| | "We are stretched so thin, and there [is] no other fund[ing]" (P14). |
| | "People are needing to figure out ways to do more with the same amount of resources, but without help" (P3). |
| | "CBD created a need for more capacity. They did a great job about the design but did nothing about the resources" (P4). |
| | "Everybody else in our training committee, it's completely volunteer basically and…everybody's just got too many things to do" (P12). |
| | "They're volunteering bottom line" (P14). |
| | "The resources, time, remuneration, people are just inadequate... It's patchwork" (P14). |
| Buy-in | "It's just not realistic for people who are unpaid volunteers, working full time as physicians to do this kind of work" (P12). |
| | "Asking our staff to take more time out to do the evaluations. It's been really challenging" (P11). |
| | "Everyone is very busy and reluctant to add more to their plate" (P12). |
| | "We talk about what's going on, what's working well, what's not working well. Why do we have to do any more work for it?" (P12). |
| | "Very hard to prove with program evaluation that CBD as a whole has improved residency education" (P13). |
| | "Wonder if there's still this much in flux. Whether it's useful to start a program evaluation right away?" (P7). |
| | "Residents [are] just too burned out from surveys because of other issues" (P6). |

## Appendix M: Phase 3 Purposeful Sampling to Recruit Program Evaluators



## Appendix N: Phase 3 Information Letter

uOttawa

Title of the study: Program Evaluation in Competency-Based Medical Education: A Mixed Methods Study


| Student Investigator: | Supervisor: |
|---|---|
| Jenna Milosek (PhD candidate)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario | Katherine A. Moreau, PhD<br>Associate Professor, Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario |

Université d'Ottawa
Faculté d'éducation

University of Ottawa
Faculty of Education

145, Jean-Jacques Lussier
Ottawa ON K1N 6N5
Canada

www.education.uOttawa.ca


Dear Program Evaluator,

You are invited to take part in Phase 3 of this three-phase research study by elaborating on your experiences with program evaluation within medical education. The purpose of this study is to explore the extent to which Competence by Design (CBD) programs are engaging in program evaluation, the ways in which these programs are engaging in program evaluation, and the reasons why they are engaging or not engaging in program evaluation. It also explores the barriers and facilitators for doing program evaluation as well as ways that CBD programs can build and support their capacities to do program evaluation and use program evaluation findings.

If you agree to participate in this phase of the study, you will be asked to take part in a one-on-one interview. The online interview will be conducted remotely. The interview will take place using Zoom at a time that is convenient for you. We will ask that you select a location for the interview that is private and convenient for you. Jenna Milosek will then conduct the interview with you from a private office. The interview will be a semi-structured conversation-style session. In the interview, you will be asked open-ended questions on how you support CBD programs in program evaluation, and how program evaluators can help these programs build their capacities to do program evaluation and use evaluation findings. The interview will take approximately **60 minutes** to complete. The interview will be audio-recorded and transcribed by Jenna Milosek for analysis. Participants will be selected on a first-come, first-served basis.

Some of the questions in the interview may make you feel uncomfortable because they ask about how you support CBD programs to engage in program evaluation. You do not have to respond to any questions that make you feel uncomfortable. You may refuse to participate in the study. Your responses, recordings and transcripts will remain confidential.

You may or may not benefit directly from this study. However, the findings from it will provide you with information on the extent and ways that CBD programs are engaging in program evaluation. It will also help you understand why programs are engaging or not engaging in program evaluation as well as the barriers and facilitators of doing program evaluation. Moreover, it will provide you with strategies for building and supporting your program's capacities to do program evaluation and use evaluation findings. To thank you for your participation in the interview phase of the study, you will receive a summary of the findings and a $10 Amazon gift card upon submission of your email address.

**If you are interested in participating in a one-on-one interview for this study, please contact Jenna Milosek by November 30, 2021 to obtain additional information about the interview.** As mentioned, the interview will be scheduled at a time that is convenient for you. Your written consent will be obtained prior to the interview. We will email you a copy of the consent form to review prior to the interview.

If you have any questions about the study, please contact Jenna Milosek or Katherine Moreau at the coordinates below. If you have any questions regarding the ethical conduct of this study, you may contact the Ethics Office (613-562-5387, room 154).

Sincerely,

| | |
|---|---|
| _____ | _____ |
| Jenna Milosek (PhD candidate)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario | Katherine Moreau, PhD (Supervisor)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario |

## Appendix O: Phase 3 Consent Form

uOttawa

Université d'Ottawa
Faculté d'éducation

University of Ottawa
Faculty of Education

Title of the study: Program Evaluation in Competency-Based Medical Education: A Mixed Methods Study

| **Student Investigator:** | **Supervisor:** |
|---|---|
| Jenna Milosek (PhD candidate)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario | Katherine A. Moreau, PhD<br>Associate Professor, Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario |

**Invitation to Participate:** I am invited to participate in a one-on-one interview for Phase 3 of this three-phase research study conducted by Jenna Milosek, PhD (Education) candidate from the University of Ottawa, under the supervision of Professor Katherine Moreau.

**Purpose of the Study:** The purpose of this study is to explore the extent to which Competence by Design (CBD) programs are engaging in program evaluation, the ways in which these programs are engaging in program evaluation, and the reasons why they are engaging or not engaging in program evaluation. It also explores the barriers and facilitators for doing program evaluation as well as ways that CBD programs can build and support their capacities to do program evaluation and use program evaluation findings.

**Participation:** My participation will consist of taking part in one one-on-one interview. The online interview will be conducted remotely. The interview may take place via Zoom at a time that is convenient for me. I will be asked to select a location for the interview that is private and convenient for me. Jenna Milosek will then conduct the interview with me from a private office. The interview will be a semi-structured conversation-style session. In the interview, I will be asked open-ended questions on how I support CBD programs in program evaluation, and how program evaluators can help these programs build their capacities to do program evaluation and use evaluation findings. The interview will take approximately **60 minutes** to complete. I understand that the interview will be audio-recorded and transcribed by Jenna Milosek for analysis. I understand that participants will be selected on a first-come, first-served basis.

**Benefits:** I may or may not benefit directly from this study. However, the findings from it will provide me with information on the extent and ways that CBD programs are engaging in program evaluation. It will also help me understand why programs are engaging or not engaging in program evaluation as well as the barriers and facilitators of doing program evaluation. Moreover, it will provide me with strategies for building and supporting CBD programs' capacities to do program evaluation and use evaluation findings. To thank me for my participation in the

145, Jean-Jacques Lussier
Ottawa ON K1N 6N5
Canada

www.education.uOttawa.ca

interview phase of the study, I will receive a summary of the findings and a $10 Amazon gift card upon submission of my email address.

**Confidentiality and anonymity:** I have received assurance from the researchers that the information I will share will remain strictly confidential. I understand that the information I provide in the interview will only be used for describing how I support CBD programs in program evaluation, and how program evaluators can help these programs build their capacities to do program evaluation and use evaluation findings. Only Jenna Milosek and her supervisor will know my identity, and I will not be asked to state my name in the interview. Any information that could potentially reveal my identity (e.g., name, specific position in an organization, years of experience) will be eliminated from the audio recording and transcript so that I cannot be identified in published reports or presentations. Thus, anonymity can be guaranteed in the dissemination of the findings. Jenna Milosek will email me a copy of my interview transcript to review and revise prior to analyses. I will have two weeks to review my transcript and share any necessary changes. Jenna Milosek will ensure that the transcript is secured in her password-protected laptop.

**Conservation of data:** The digital audio recording of the interview will be downloaded and erased from the audio-recorder immediately after the interview. All data will be stored on a password-protected laptop. Only the members of the above-mentioned research team will have access to the data. Data will be conserved for five years after the publication of research findings. After this time, data will be deleted, shredded, and appropriately discarded.

**Voluntary Participation:** I am under no obligation to participate and if I choose to participate, I can withdraw from the study at any time and/or refuse to answer any questions without suffering any negative consequences. If I choose to withdraw, all data gathered until the time of withdrawal will be deleted, destroyed, and not included in any publications or presentations.

**Acceptance:** I _____, agree to participate in the above-mentioned research study conducted by Jenna Milosek PhD (Education) candidate from the Faculty of Education, University of Ottawa under the supervision of Professor Katherine Moreau.

If I have any questions about the study, I may contact Jenna Milosek or Katherine Moreau at:

| | |
|---|---|
| _____ | _____ |
| Jenna Milosek (PhD candidate)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario | Katherine Moreau, PhD (Supervisor)<br>Faculty of Education<br>University of Ottawa<br>Ottawa, Ontario |

If I have any questions regarding the ethical conduct of this study, I may contact the Ethics Office at (613-562-5387, room 154).

I will save or print a copy of this form for my personal records.

Participant's signature: _____ Date: _____

Researcher's signature: _____ Date: _____

## Appendix P: Phase 3 Interview Guide

Phase 3 Program Evaluator Interview Guide – Using Program Evaluation in CBD

***For use with program evaluators who use program evaluation in CBD context***

DATE: _____
TIME: _____
PARTICIPANT ID: _____
INTERVIEWER: _____

To begin, I would like to confirm some information to ensure accurate context for the interview.

1. Are you involved in the evaluation of CBD programs? _____
   - If 'Yes,' continue with this interview guide.
   - If 'No,' thank them for their time and end the interview.

2. Tell me a little bit about your experience as a program evaluator?
   - How many years of experience do you have as an evaluator?
   - How long have you been involved in the evaluation of medical education programs?

***Research Question 5: How are program evaluators currently supporting Competence by Design (CBD) Canadian medical education residency programs in program evaluation?***

3. How do you and/or your team support the evaluation of CBD programs?
   a. What is/was your role as the program evaluator?
   b. Who have you worked with in your role as an evaluator on your most recent CBD program evaluation?
   c. Can you tell me about your experience designing the program evaluation?
      - Can you tell me about the evaluation approach or strategy was used in designing the evaluation?
   d. Can you tell me about your experience implementing the program evaluation?
      - Can you tell me about the evaluation process?
      - Can you tell me about the evaluation activities?
      - How is/was data collected to support the evaluation?
      - What types of data are/were collected to support the evaluation?
      - How are data analyzed?
   e. How are/were the evaluation findings disseminated/reported?
   f. What happened after the program evaluation to the evaluation findings?
4. Why are you and/or your team supporting CBD programs in program evaluation?
      - What helps you and/or your team do program evaluation with CBD programs?
5. Are there any barriers or challenges for you and/or your team in doing program evaluation with CBD programs?
      - If so, what are these barriers or challenges?
      - Do you have suggestions to reduce or minimize these barriers or challenges?

***Research Question 6 (1st half): How can program evaluators help CBD programs build their capacities to do program evaluation?***

6. Do you and/or your team support stakeholder participation in the CBD program evaluation?
    - If so - What training have you and/or your team provided to support CBD program stakeholders in program evaluation?
    - If not - Are you and/or your team interested in involving stakeholders in the CBD program evaluation?
7. Do you and/or your team provide evaluation capacity building activities to CBD programs so that they can continue to do program evaluations without an evaluator?
    - If so - How and why?
    - If not - Why?
8. Do the availability of resources (e.g., human, financial, time) affect your ability and/or your team's abilities to help CBD program stakeholders do CBD program evaluation?
    - If so - How?
9. In your opinion, how can program evaluators help CBD programs build their capacities to do program evaluation?
10. Based on your recent program evaluation scenario or experience, if you could make changes in the way that you support CBD programs in evaluation, what would these changes be?

***Research Question 6 (2nd half): How can program evaluators help CBD programs build their capacities to use evaluation findings?***

11. Based on your recent program evaluation scenario, please share how program evaluators can help CBD programs build their capacities to use evaluation findings?
12. Are you concerned about if or how program stakeholders use their program evaluation findings?
    - If so, how do you and/or your team help and/or encourage program stakeholders to use their program evaluation findings?
13. In your opinion, how can program evaluators help CBD programs build their capacities to use program evaluation findings?
14. What are some of the barriers or challenges that may prevent you and/or your team in helping CBD programs to build their capacities to use evaluation findings?
    - Do you have suggestions on how you and/or your team could reduce the challenges or barriers to help CBD programs build their capacities to use evaluation findings?

15. Would you be interested in receiving a summary of the findings?
    O Yes O No
    *Is this still your email address?* _____

16. Thank you for participating in this follow-up interview. You will be compensated with a $10 Amazon gift card for your participation in the follow-up interview.
    *Confirm email address:* _____

17. I will email you a transcript of this interview for you to review and make any changes in your responses. I will email you in the next day or two. Please review it and make any changes within two weeks.

**Thank you for your participation!**

## Appendix Q: Phase 3 Final NVivo Codes

**Program evaluators are supporting CBD programs in program evaluation by:**

- Responding in a reactive way as a temporary and external evaluation consultant
- Collaborating with interested stakeholders as internal evaluation experts from a university
- Leading an evaluation approach that is manageable for program directors with limited time
- Using methods that complement the PGME environment
- Using evaluation processes that lead to the publication of findings
- Offering professional development sessions about program evaluation to support accreditation activities

**Program evaluators can help CBD programs build their capacities to do program evaluation by:**

- Using a participatory evaluation approach
- Offering evaluation resources and expertise as guidance
- Leveraging existing data
- Advocating for clearer program evaluation policy expectations from the RCPSC.

**Program evaluators can help CBD programs build their capacities to use evaluation findings by:**

- Recommending that program evaluation is included at the beginning of:
    - Program design
    - Implementation
- Including the stakeholders who will be using the findings in the evaluation itself
- Focusing stakeholder questions on meaningful program areas over which they hold authority to make changes
- Encouraging the use of program evaluation approaches that are appropriate to the CBD implementation context
- Interpreting the findings:
    - Alongside the stakeholders
    - Co-develop a manageable follow-up plan
- Encouraging programs to share findings which establishes an accountability cycle to make changes based on findings

**Appendix R: Phase 3 Sample Data Analysis Matrix**

| Factor | Quotations |
|---|---|
| Program evaluators are supporting CBD programs in program evaluation by: | |
| Responding in a reactive way as a temporary and external evaluation consultant | "Consultant role, it's much more of 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).<br>"Is the way that [they] don't favour . . .—when they become interested in program evaluation after the fact" (E2).<br>"Already been developed, it's already been implemented. And then they sort of want to collect some data on how things are going after the fact" (E2).<br>"All the work and resources get put in and then a report gets shoved and hidden in a drawer and collects dust" (E2).<br>"Not saying we want all you programs to do program evaluation, and here are some resources and supports for everybody to do it" (E2).<br>"Minimum requirements of what we want you to evaluate. None of that of this occurring" (E2).<br>"It absolutely is up to the individual program directors to take on the responsibility of program evaluation. And I would even say to even decide if they're going to do any program evaluation" (E2).<br>"Are so overwhelmed by all the changes and everything that they need to do just in their day-to-day that the thought of adding integrating something else within it really overwhelms their mind" (E2).<br>"Not come across an RFP [request for proposal] for that. So that's why [many evaluators are] not doing it because nobody has invited [them] to do it" (E1). |
| Collaborating with interested stakeholders as internal evaluation experts from a university | "Feel like [program directors are] not really committed to evaluation, then it's not something [they would] want to get engaged in" (E1).<br>"You can't really do good evaluation, unless you've got interest from the organization" (E1). |

"First real carrot was [program directors] genuinely wanted to improve" (E8).

"Have been engaged in program evaluation with Competence by Design [that] has been much more about a collaborative fashion" (E2).

"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).

"Their program director, their program administrator and their residents for a focus group and surveyed faculty and residents" (E10).

"The hospital supervisors, academic advisors and Competence Committee members, [and] the clinical supervisors" (E8).

"Resident [who] was a superstar who was . . . pushing and leading [the evaluation] and [she] and [her program director] were going back and forth" (E5).

"Somebody says how about a patient survey? Okay, do we have any existing patient surveys? No. Okay. We need to develop one" (E5).

"Reach out to us telling us we need to do this and that and we think of evaluation framework and how we're going to approach the concerns. And then we'll provide guidance along the way" (E4).

"An internal evaluation unit . . . in medical education . . . in the undergraduate environment, [because] it's a lot more uniform . . . it's just one learning environment survey for everybody" (E3).

"A multidisciplinary group . . . responsible for providing program evaluation" (E6).

"Don't know how much [post graduate CBD evaluation] exactly happened . . . we weren't doing anything" (E6).

"Also hearing . . . the Royal College as one source . . . were trying to provide some support to do evaluation or monitoring or gathering data around [postgraduate medical education] . . . so, [they weren't] feeling quite so guilty" (E6).

"Where they found us our [evaluation] services really, really useful is that we're able to kind of get into the smaller programs without the risk of kind of the anonymity or confidentiality of the residents" (E3).

"For [smaller programs], they're finding the value to having an internal/arm's length internal unit to do this work" (E3).

"The idea that we were just waiting for them to come and seek help. I don't know if this is helping or not. Because we're not getting a lot of requests" (E4).

"The Royal College supporting program evaluation to a great extent [at] . . . summits and many presentations about program evaluation . . . [about] where to get help on program evaluation" (E4).

"An internal unit that services the entire undergraduate and postgraduate education and as well as health professions education. So, [they] are of course limited in resources" (E3).

"Senior leadership . . . [who] directs and gives priorities [to] program areas that we should be working in" (E3).

"They don't want to use [their] services for some programs and not others, [thus they are] still in the process of figuring out [how] to support the entire residency programs and in a systematic way" (E3).

"Postgrad, because it is so decentralized and all the programs do their own things, [programs] are all at very different stages of implementing CBD" (E3).

"Tried . . . doing evaluation work that will meet the common good or broad needs to inform planning and decision making for in postgrad generally" (E6).

"The postgrad deans are in a place where they're trying to figure out exactly how to use [them] to get the most bang for their buck" (E3).

"An overall common framework for evaluation—would that meet needs? Or would you have to come up with tailored evaluation strategies for each of the programs?" (E6).