

CBME Program Evaluation Summit

Opening of the Summit

Dr. Andrew Hall, Chair of the CBME Program Evaluation Summit, opened the day by welcoming everyone and sharing our goals for the Summit.

- Discuss how to evaluate CBME implementation
- Understand the strengths and challenges of implementation efforts
- Share lessons learned and find out about others
- Improve and adapt our CBME programs
- Find collaborators and build a community around CBME evaluation
- Understand if culture change is happening within our program or institution
- Figure out if CBME is worth all of this effort

The Royal College vision for program evaluation involves program evaluation from outside the College – it is unrealistic for the Royal College evaluation team to be able to do all the work to drive the evaluation, and it is imperative that program evaluation work also come from outside to avoid bias.

The Royal College program evaluation framework has three pillars of the evaluation that will be referenced throughout the day, and helped form the themes of the day.

- Readiness to Implement
- Fidelity and Integrity of Implementation
- Outcomes

Introduction to Program Evaluation and Priority Questions

Dr. Elaine Van Melle and Dr. Tim Dalseg led the group through a discussion on what is program evaluation, and priority evaluation questions.

They reviewed the cycle of program development, and then posed the questions:

- What is your number one program evaluation question?
- What are our priority program evaluation questions?

How does program evaluation compare to research, quality improvement, and accreditation?

- Research **proves**, quality improvement **improves**, accreditation **approves**, and program evaluation **moves**.
- Quality improvement is about the **best practice**, while program evaluation is about the **next practice**.
 - We continue to learn and move forward.

Program evaluation questions are

- Relevant to program stakeholders
- Connect/consider process and outcomes
- Provide timely information for decision-making (technical report)
- Move a program forward (NEXT practice)

Small group breakout

The Summit moved to small group breakouts, where attendees wrote their number one program evaluation question, then discussed similarities and differences and additional evaluation questions with those at their table. The whole Summit then discussed common themes.

- Culture
- Different levels of influence (institutional, specialty, etc.) and which is most important?
 - There are sometimes differences in how specialties vs. individual programs are doing on their implementation
- Tensions (cost vs. quality, standards, etc.)
- Lack of leadership
- Amount of flexibility at the local level
 - Programs have different resources, making it difficult to follow guidelines
- Process vs. outcome
- Do residents choose EPAs based on positive clinical performance?
 - What is the comparison between resident and faculty triggered EPAs? Does this make a difference in Competence Committee progression?
- Are progression decisions based on potentially missing data?
- How do EPAs and WBAs cover medical and non-medical expert – do they?
 - Should we assess readiness for practice just based on EPAs?

Takeaway: Think about your individual priority question: how will you establish the next practice?

Readiness and Fidelity and Integrity of Implementation

Dr. Andrew Hall and Dr. Warren Cheung reviewed Readiness to Implement and Fidelity and Integrity of Implementation.

Readiness to Implement

Readiness to implement is important because implementation affects outcomes, and an organizations' readiness for change is an important precursor to successful implementation. Therefore, it is important to build an understanding of the factors that influence the capacity to successfully implement CBD.

Readiness to implement involves both an organization's **resolve** and **capacity** to implement an innovation. At the national level, the examination of readiness to implement is focused on the R =

MC² framework; this includes motivation, general capacity, and innovation specific capacity as interactive components of readiness.

Another important aspect of readiness is modifiable vs. non-modifiable factors, and how modifiable factors can be optimized to ensure successful implementation.

Fidelity and Integrity of Implementation

Fidelity of implementation is the extent to which critical components of CBD are present in a program, and **integrity of implementation** is the extent to which implementation embodies key qualities of CBD.

Fidelity and integrity of implementation are important because they allow us to understand the influence of local contexts, and determine if a lack of impact is due to implementation factors or inadequacies in program theory.

CBD implementation is flexible, and each local context will adapt in different ways. Fidelity and integrity allow us to examine if a program's implementation of CBD include critical components of CBME, and if a program's CBD program embodies the key qualities of CBME.

An attendee commented that fidelity and flexibility seem like opposite components. Dr. Hall clarified that there are minimal components of CBME that need to be present in a CBD program, but how they are operationalized may be different at different sites. Integrity can help measure if this operationalization is true to the integrity of CBME.

Poster sessions and small group discussions

After the poster session, tables were split into 6 topics, and three questions were posed.

- What specific aspects of a training program would you measure relating to this component of CBD?
- What are strategies that can/should/are being used to evaluate your component of CBD?
- How could you link this fidelity evaluation to subsequent outcomes?

Readiness to Implement Table

The discussion involved

- Different conceptualizations of readiness
 - Different levels, such as individual, program, discipline, and institutional, and how readiness at one level may depend on another
- Key readiness structures that need to be in place
 - For example, if an electronic platform isn't in place, there likely won't be buy in
- Whether readiness indicates a minimal level or an optimal level?
 - Additionally, when does readiness end?

Coaching and Individualized Stage Based Learning Table

The discussion involved

- How to best define a coach
 - This is an issue, a coach needs to be defined
 - It is hard to evaluate if there is not shared terminology and mental models
- How to best select and develop a coach. Is there a framework?
- How to make sure there is a coach for each resident
- Do residents really develop a coaching relationship? If so:
 - Is this valuable and helpful?
 - How much time does it take to develop this?
 - Is there a conflict between coaching and assessment?

Workplace Based EPA Assessment and Direct Observation Table

The discussion involved

- How many WBAs are required, and how different numbers are decided on.
- Context, and how things will need to look different in different places
 - How to make CBD work for all disciplines, all sites.
- Is there a difference in assessment based on who triggers them (resident vs. faculty) and if residents should trigger assessments
- What competencies can be measured inside and outside of the workplace
- Creating a WBA system that captures the difficulties in implementation
 - Does direct observation create workflow issues? Time demand, cognitive demands
- Practicality (number to demonstrate competence, best practices for timing and nature) and theoretical and philosophical (holistic care, resident motivation to learn and improve)
- The value of a number scale vs. qualitative descriptions
 - How would we evaluate narrative feedback?
 - Is direct observation high yield? People will put in the time if they see it as high yield.

Programmatic Assessment Table

The discussion involved

- What is good assessment, and what does it look like?
- How to avoid simply checking boxes, without thinking about fidelity and integrity
- How to reflect that narrative feedback is better
- To measure programmatic assessment, measure variability in assessments, assessors, count narrative, types of assessment, and different points of assessment
- To link to outcomes, look at the intended vs actual implementation – curriculum maps, content analysis, and portfolios
- Think of how a specific outcome can link to fidelity. For example, does fidelity in a specific aspect of programmatic assessment correlate with ease of making decisions?
- Ask why programs chose various types of assessment

Competence Committees and Progression Decisions Table

The discussion involved

- Competence Committees appear to have consistency, but once you get into the weeds and listen to conversations about progression, there are differences in how decisions are made
- Some people still want to have something that says someone passed, such as an ITER. However, this takes the decision away from the Competence Committee and puts it back in the hands of the person who filled out that document.
- Program directors that sit on the Competence Committee take on a different role
- How are decisions from the Competence Committee transmitted back to residents for learning plans? What are the challenges involved in this?
- What is the best way to distill down the volume of data and present it to Competence Committee members?

Next Steps Table

The discussion involved

- The difference between the date of implementation and integrity of implementation
- Helping programs navigate an evolutionary and deep systems change, and leveraging national standards without being too prescriptive
 - How do you ensure implementation is true to the vision without being too prescriptive?
- Resources, putting mechanics in motion is a challenge

Outcomes

Dr. Lara Cooke opened the afternoon session by discussing outcomes. She presented various models and frameworks that can be used to look at outcomes, such Kirkpatrick, a program model, time based, and core components based.

She then invited the group to think of the promise of CBD, and think of how we can measure outcomes to determine if we are meeting the promise of CBD. We need to look critically at the promise of CBD.

Finally, she discussed the “holy grail” of outcomes – better patient and community outcomes.

Poster Session and Small Group Breakouts

The group broke for a poster session, and came back to tables that were split by topic. The topics were chosen from common topics that arose at previous meetings, and are based on the core components, and short, medium, and long-term, and speak to various (Kirkpatrick, Guskey & Moore) evaluation frameworks. Each table had three questions to think about:

- What else do we need to know in order to clearly define this domain?
- How should this domain be evaluated?
- What is needed/what are the barriers to evaluating this locally, provincially, and nationally?

Unanticipated outcomes Table

The discussion involved

- The challenge of defining what outcomes were intended. Everyone had a different idea of intended outcomes.
- One idea for unexpected outcomes is that they are things not captured by EPAs, such as the hidden curriculum.
- Unintended outcomes can be evaluated qualitatively through discovery (ethnographic, etc.) and through talking to people in an inductive approach.
 - Probing where there are misalignments or friction points may highlight the unexpected.
- They could be accessed through a retrospective mode, or a systematic, continuous evaluation. Having something standardized could capture them.
- Some barriers to this are protecting the safety of patients, residents, programs, etc.
 - Must have safety in order for people to tell the truth
 - Access to documents is a challenge, and that there is no common database or sharing through platforms.

Patient Outcomes Table

The discussion involved

- Patient outcomes could start with advocacy. For example, better access to health systems data.
- Patient outcomes could be specialty specific.
- To evaluate this, could examine resident sensitive quality indicators, complaints data (such as communication issues, professionalism).
- Consider the role medical education plays on what society is looking at.
- Patients should be involved at every level.
- Patient outcomes are beyond just medical education, how should we link with what is already established.

Changing Role of the Preceptor/Coach/Educator Table

The discussion involved

- The tensions involved in this dimension
 - There is a lot of tension between someone having multiple roles at the same time, such as supervisor, patient safety advocate, assessor, coach, mentor, etc. The concept of coach competes with other roles.
 - It is difficult mentally and cognitively to deliver on all fronts
 - This may add the need for more resources.
- The perception of the coach may be different between faculty and resident – the coaching may be lost through the nature of the assessment piece for the resident.
- Academic advisors could be a layer in a coaching mentorship.

- The role of coach in medicine is difficult because most coaches normally coach during practice, but in medicine every day is game day.
- Need to reach consensus on the definition of a coach, and understand this more fully.
- Different specialties likely approach this differently – what does it look like in the day-to-day of each speciality? How is each role integrated?

Earlier Identification of Residents in Difficulty Table

The discussion involved

- The ways in which we define a learner in difficulty
 - When is a learner premature vs. delayed?
 - When do scores of 2 or 3 become problematic? When is it the expected trajectory vs. an issue?
 - When do short programs need to make these decisions?
- The challenge is having a pre-post approach. The pre is not detailed or consistent, and is not necessarily an accurate representation.
 - Can look at this in terms of early identification prospectively, but this is a hard sell design – how do we show it as the truth?

Transformation of the Culture of Assessment in Residency Table

The discussion involved

- Knowing if a culture of assessment is moving towards a CBD vision. The group thought of some markers to use to show a program has evolved.
 - Coaching mindset
 - Developmental view of assessment
 - Psychological safety
 - More observation, especially direct observation
 - Finding residents in difficulty earlier
 - Better quality coaching comments, actionable
 - Timeliness of assessments (triggered and actioned in real time)
 - Capturing non-EPA data
 - Less gaming, more natural variability. Slopes of developmental markers
- Barriers to evaluating this are having a shared definition, operationalizing, and having access to data.

Residents' Readiness to Transition to Practice Table

The discussion involved

- The barrier of the difficulty in articulating nuances of independent practice
- The table developed a framework for transition to practice
 - Medical expert piece still needs to be developed and nurtured
 - Autonomy and how to get at this
 - Leadership and management
 - Administrative piece (billing, etc.)

- Professionalism, collaboration, communication
- Teaching
- Physician wellness
- To evaluate this, we can use examinations, 360 evaluations, Transition to Practice EPAs, self-reports on confidence, etc.
- We need access to data, including data on recent graduates in practice in order to compare.

Dr. Lara Cooke closed the discussion by summarizing some common themes from the discussion.

- Trust, especially for learners in difficulty
- Pre-post: Comparators will be a question mark for many of these
- Access to data

A last thought from Dr. Elaine Van Melle was that we need to detach from outcomes to a certain extent. If we truly believe the changes we are making are worth spending time on, we have to have some faith that outcomes will follow.

Closing Plenary Panel – Exemplary Program Evaluation

Dr. Deena Hamza from the College of Family Physicians of Canada (CFPC), Dr. Stanley Hamstra from the Accreditation Council for Graduate Medical Education (ACGME), and Dr. Warren Cheung and Dr. Tim Dalseg from the Royal College of Physicians and Surgeons (RC) presented program evaluation projects from their respective organization.

Improvement-Oriented Evaluation of CBME – The College of Family Physicians of Canada

The CFPC conducted a program evaluation of the Triple C Curriculum. This program evaluation was a utilization focused evaluation, which involved a deliberate decision to focus on collecting data that was useful.

A key factor of this study was collaboration. This involved engaging the intended users at the beginning, including in the program evaluation plans. They then followed up with the intended users and presented the data in a report to facilitate use and expand influence. This ensured that they were accountable and could learn from the evaluation and improve on the intervention.

The program evaluation was a combination of process and outcome evaluation, exploring social processes and mechanisms during implementation and prospectively drawing a bridge to outcomes. Outcome evaluation then explores the progress of CBME and the desired results.

The type of evaluation (utilization focused) feeds into and aligns with CQI. This helped to develop a program theory or theory of change.

- Clearly defines the problems/challenges CBME was meant to address
- Illustrated how the shift to CBME was to be successful
- Defined the intended impacts

- Systematically mapped all of the factors that contributed to outcomes that are expected to lead those impacts.

The study had some assumptions.

- Assumption 1: Policies and support offered will enable successful adoption of Triple C
- Assumption 2: Update will vary depending on external factors and internal factors
- Assumption 3: If family medicine trainees experience Triple C, graduates will choose to practice comprehensive family medicine, will choose to work in diverse communities, and will be able to self-assess and address ongoing learning needs.

Dr. Hamza did a deep dive into Assumption 1.

- Data sources included the Residency Program Implementation Profile and the Qualitative Understanding and Evaluation Study of Triple C Study.
- A lot of uptake of Triple C occurred before the accreditation standards required it.
- Collaborative co-creation with stakeholders supported adoption.
- Effective communication with all program leaders was imperative.
 - Including the PG Deans, who were not engaged in the beginning.
- Longitudinal support from the CFPC is needed for late adopters.

Key takeaways

- Co-creation and collaboration of evaluation is important
- Two guiding principles
 - Collective action (collaborative, co-creation, collective goal, decentralized authority)
 - Accountable: Learn and Improve (data used to inform upcoming processes, encouraged engagement in evaluation and research, participants had evidence they contributed to the growth of Triple C).
- Program theory approach allows for new changes and theories that facilitate the advancement of Triple C.

Realizing the Promise of CBME with Milestones – Accreditation Council for Graduate Medical Education

The ACGME collects milestones data every 6 months. This data represents an opportunity to engage in an ongoing CQI process. This isn't necessarily a program evaluation, but they found many indirect indicators of impact.

One example of this is "Change in Educational Practice – The Neurosurgery Story". With milestones, the ACGME can monitor progression of Milestone achievement in multiple competency categories within each speciality. Level 4 is the target for graduation, but some residents don't reach level 4 and still graduate.

In neurosurgery, many residents were not achieving a level 4 in some sub-competencies. The ACGME asked neurosurgery "why is this? Is this okay?" Neurosurgery reviewed their milestones and found that some of the sub-competencies listed were more appropriate for fellowships, or

depended on training site. As a result, they revised the Milestones language, content, and structure, and created Milestones 2.0.

Milestones 1.0 did not have a model; but with the data and ongoing CQI, they can be revised. It is important to have co-creation in creating and revising the milestones, and to interpret the meaning of the data.

CBD Program Evaluation: A review of the Readiness to Implement and Pulse Check Studies – The Royal College of Physicians and Surgeons of Canada

Readiness to Implement Checklist

The CBD Readiness to Implement Checklist was conducted to help address the goal of fostering a successful implementation of CBD. It followed the $R = MC^2$ framework, touching on motivation, general capacity, and innovation specific capacity. These are interactive components that come together to help measure readiness.

The study aimed to assess readiness to implement CBD in the 2019 cohort, to identify challenges and areas of success in the lead up to implementation, and provide programs with a resource checklist to guide their preparation.

A survey was sent to program directors and/or program CBME leads of the 2019 launch disciplines in June, just prior to launch. The survey addressed motivation, general capacity, and innovation specific capacity.

In terms of *motivation* respondents felt that a successful implementation of CBD was a priority, but questioned if CBD is a move in the right direction and if the implementation of CBD is a manageable task.

In terms of *general capacity* respondents felt that leaders in their program were supportive of change, and that their program was receptive to change. However, fewer programs had previous experience with change or adequate support staff.

Respondents were then asked if they had completed various pre-implementation tasks. On average, programs had completed 72% of pre-implementation tasks. Typically, tasks that were administrative in nature, such as identifying Competence Committee members and creating a curriculum map, were more likely to have been completed than activities relating to on the ground implementation, such as teaching faculty how to incorporate direct observation and teaching into workflow.

Only approximately 60% of programs had adequate administrative support in place and trained, and less than 50% of programs had engaged with off-service disciplines.

Moving forward, the team wants to think how CBD implementation can be made more manageable. This includes identifying ways of addressing readiness gaps, finding patterns of readiness to learn

what factors are associated with successful implementation, and correlating the Readiness Checklist with the Pulse Check and outcomes studies.

Pulse Check

The Pulse Check was conducted to monitor the status of implementation of CBD across the system, and to gain an understanding of the challenges and opportunities to improve implementation, as well as early outcomes and advice for moving forward. The study focused primarily on fidelity and integrity of implementation.

The study involved a survey of program directors or program CBME leads of the 2017 and 2018 CBD launch disciplines, and a follow up interview with a subset of the participants. The survey asked about overall implementation, status of the implementation of key features of CBD, faculty development, and benefits, challenges, and advice for moving forward. The interview dove deeper into their experience.

The survey used innovation configuration mapping to create the scales for the key components of CBD. Each point on the scale had a unique description pertaining to that component so programs could identify how far along they were on the scale. The scale for each component ranged from not yet implemented to full implementation.

Overall, programs felt their CBD implementation was going okay to well. Programs were working towards full implementation on most key components of CBD. Of note, 21% of programs were not yet or sometimes using an electronic platform and 33% were not yet or rarely using individualized learning plans. These are areas to examine further.

Common benefits included higher quality and quantity of feedback, more objective resident assessment, early identification of struggling residents, and better faculty and resident engagement. Common challenges included the time involved, challenges with EPAs (completion, amount), faculty and resident buy-in, and the functionality of the electronic platform.

The team identified some draft recommendations for moving forward.

- Provide clear, easy to access information and resources for programs.
- Encourage and facilitate the sharing of best practices.
- Learn from past challenges.
- Share information and early outcomes of CBD, monitor negative outcomes.
- Improve electronic platforms.

The Royal College presentation closed with a look at upcoming projects, showing that there will be a series of repeated projects and that the creation of more projects is in the works.

Discussion

Summit participants raised the idea of targeting non-responders. This is something the CBD Program Evaluation Operations team can look into, and other evaluation strategies, such as rapid evaluation, will be more targeted.

Another participant raised the idea of program evaluation fatigue – many projects are going on across many sites. We need to ensure we aren't duplicating things to minimize fatigue. There is a need for novel studies.

Reflections on CBME Program Evaluation

Dr. Jonathan Sherbino closed the summit with his reflections on the current state of CBME Program Evaluation.

He reflected on four themes from the Summit, and noted that attendees of this Summit should serve as multipliers, and expand our conversation beyond these walls.

1. Program evaluation is necessary.

- It is difficult to rally educational scientists with program evaluation, but we hope to engage them. We must have conversations about the utility of program evaluation and how it plays a role in a theory driven discipline.
- Health services research can be an interesting and useful paradigm to look at.
- What conversations must we have at home to engage in program evaluation?

2. The importance of data

- There are good studies that truncate at the program or institutional level. There is now an opportunity for all Faculties of Medicine to find a third party steward of data.
- To ask big questions it is necessary to move past the current barriers and understand what a common data sharing agreement looks like.

3. It is time to think about new and emerging analyses

- Should engage experts in other domains to co-create scholarship, current methods may not answer the hard questions we are wrestling with.

4. We lack a shared language

- Is it time to have a "CBME extension" so as innovations are reported, others can replicate and compare?