Post MD Education
UNIVERSITY OF TORONTO

Best Practices in Evaluation and Assessment (BPEA)

Program Evaluation and Monitoring of Competence by Design (CBD)

Martin Schreiber
Director, Preclerkship Faculty of Medicine, University of Toronto
Associate Professor, Faculty of Medicine, University of Toronto
Staff, St. Michael's Hospital – Division of Nephrology

Andrew Steele
Program Director, Adult Critical Care Medicine Residency, Faculty of Medicine, University of Toronto
8. Assistant Professor, Department of Anaesthesia, Interdepartmental Division of Critical Care,
Faculty of Medicine, University of Toronto
Staff Intensivist and Staff Anaesthesiologist, UHN, Toronto General Hospital

). **Patrick Fleming MD, MSc**
1. Division of Dermatology, Department of Medicine, University of Toronto
2. Chief Dermatology Resident

3. **Caroline Abrahams MPA**
4. Director, Policy, Analysis, and Systems Management,
5. Faculty of Medicine, University of Toronto

UT CBME / CBD

Best Practices in Evaluation and Assessment (BPEA)
Program Evaluation and Monitoring of Assessment for Competence by Design (CBD)

Martin Schreiber, Andrew Steele, Patrick Fleming, Caroline Abrahams

1. Executive Summary

The implementation of Competency-Based Medical Education (CBME) for residency education at the University of Toronto represents a major change in the way physicians will learn and train, and in the way they will be observed, supervised, and assessed. Given the significance of this change, it is critical that careful attention is paid to the program evaluation of the initiative. Program evaluation and monitoring of CBME will ensure there are ongoing and iterative adjustments to implementation approaches as required, and will ensure that programs are accountable by responding to lessons learned. While there is not a great deal of literature on program evaluation specifically for CBME, the working group undertook a review of 22 relevant articles and reports to identify key implications for Postgraduate Medical Education (PGME) at University of Toronto (U of T) as follows:

- Program evaluation for CBME must be contextual, flexible (utilization-based), and **developed at the same time as learner assessment tools**, in order to maximize relevance and stakeholder buy-in of this change. Evaluation tools should not be developed as an "afterthought." Feedback for program directors, faculty, and administrators on the implementation of CBME is as important as the assessment tools developed for learners. It is this feedback, through program evaluation, that will sustain the CBME change initiative.
- Evaluation measures and processes must be balanced in volume and length so as not to overshadow the learning and education experience, yet reflect the complexity of the delivery of patient care. It is necessary to start developing program evaluation tools at the same time as learner assessment tools/e-portfolios.
- PGME at U of T has recently redeveloped evaluation tools, such as the Resident Assessment of Teacher Effectiveness (RATE) and the Rotation and Educational Site Evaluation (RES), that could be adapted for the CBME context. These tools currently exist in an on-line format in the Postgraduate Web Evaluation and Registration (POWER) system and could be adapted to reflect features of CBME through mobile capability in POWER or other online systems.

2. Background

Many individuals, organizations, and stakeholders have invested significant amounts of time and resources in the shift from traditional time-based residency training to the new model of

Competency-Based Medical Education (CBME). The shift is based on overwhelming evidence and literature that supports a model of physician training that emphasizes mastery of skills and knowledge. In an effort to support the implementation of this new paradigm for training at the University of Toronto (U of T), a working group was established to consider Best Practices in Evaluation and Assessment (BPEA) for CBME at U of T. The working group comprises Program Directors, experts in evaluation and assessment, residents, and PGME staff. In order to complete its work, the working group broke into nine small groups to pursue a closer examination of several implementation considerations related to evaluation and assessment. This paper represents the work of the group examining Program Evaluation and Monitoring of Assessment in CBME.

3. Methodology

Working group members met initially by teleconference to discuss the topic, relevant considerations and issues, and how best to approach the task. During the initial teleconference, 22 articles and reports related to Competency-Based Medical Education, as well as Educational Program Evaluation and Assessment, were identified for review. Articles were identified as part of prior research conducted by the PGME Research and Education Unit. In addition, the U of T PGME reports on Best Practices in Teacher Evaluation and Best Practices in Rotation Evaluations were reviewed as part of the group exercise, along with a recent article on medical student course evaluations. The articles were divided up for review among the four group members.

Group members prepared summaries of each relevant article, highlighting the key findings for program evaluation, and the implications of the key findings as they related to the implementation of CBME at U of T. All the summaries were reviewed and synthesized for presentation to the BPEA Working Group and for the final report.

4. Results and Discussion

Several articles offered general critiques of CBME and general approaches to course and program evaluation in education, which were useful but not directly relevant to the topic under consideration. Based on the review of articles to date, several key themes have emerged.

4.1. Definitions

A number of articles highlighted the importance of distinguishing between “evaluation,” generally understood as the process of obtaining information about a course or a program, and “assessment,” which addresses the measurement of learner performance (Goldie 2006).

4.2. Tools and Approaches for Assessment and Evaluation

Several articles highlighted a range of program evaluation models. As part of the models, a number of useful tools and approaches have been developed, including the eight conditions of

Entrustable Professional Activities (EPAs) for CBME implementation by Program Directors and Committees, broad categories of Program Evaluation, including utilization-focused evaluation that is particularly suitable for PGME (Vasser, Wheeler et al. 2010), and a checklist for the development of assessment tools. There is a trend toward the use of new concepts and acronyms tailor-made for CBME, including "STARs" – Statements of Awarded Responsibility for successful completion of EPAs {ten Cate, 2007 #216}.

4.3. Cautions and Implications

Much of the literature contained cautions with respect to program evaluation in general and specifically related to CBME. First was the consideration of varying perspectives in the implementation of CBME at U of T in relation to the theory, ethics, and bias that underlie program evaluation. CBME relies heavily on measurement, with related checklists and documents, and we must ensure that efforts to "measure" do not overshadow the need to learn and educate as part of the educational process. Measurement should not take precedence over necessary activities and interactions that occur in the education of future physicians. The impact of measurement activities must be carefully reviewed to preclude unintended consequences or counterproductive efforts in the educational goals of residency programs. In the early stages of developing assessment frameworks, programs will be required to create their own measures and benchmarks. Attention must be paid to standard-setting, procedures for providing feedback, and supporting mastery learning. It is important to design questions/evaluations about teachers and learning environments that residents are in a position to answer.

In particular, it was noted that many models of program evaluation in education and health-care settings appear cumbersome, and their metrics may not capture the complexity of delivery of patient care. Designing appropriate and effective program evaluation, and monitoring that evaluation, must integrate with the many priorities involved in delivering patient care, as well as providing an educational experience. It must also not detract from important positive interprofessional relationships in a healthcare environment.

4.4. Key Findings

In reviewing and synthesizing the literature, several key findings emerged that were relevant to the implementation of CBME at U of T for residency training.

4.4.1. Related to program evaluation

- There is a need to take a systematic yet customized approach (program by program) to designing and implementing a program evaluation initiative for CBME.
- We must make sure we are collecting meaningful and relevant data. It will be important to adhere to basic measurement principles, with high levels of validity and reliability.
- "Utilization-Focused" program evaluation is an individualized approach that includes principal stakeholders in the decision-making process. In this case, it will be important to include learners and faculty, as the input of both of these groups is critical to developing buy-in for the CBME assessment strategy.

4.4.2. Related to monitoring of assessment

- Use checklists for development of assessment tools at the beginning of course development – not as an afterthought.
- Attention must be paid to standard-setting, and to procedures for providing feedback and supporting mastery learning.

6. References

Goldie, J. (2006). "Evaluating educational programmes AMEE Guide No. 29." Medical Teacher **28**(3): 210-224.

Vasser, M., D. L. Wheeler, M. Davidson and J. Franklin (2010). "Program Evaluation in Medical Education: An Overview of the Utilization-focused Approach." Journal of Educational Evaluation in the Health Professions **7**(1): DOI: 10.3352/jeehp.2010.3357.3351.

7. Additional References

Asch, D. A., A. Epstein and S. Nicholson (2007). "Evaluating Medical Training Programs by the Quality of Care Delivered by Their Alumni." JAMA **298**(9): 1049-1051.

Bland, C. J. (2000). "Curricular change in medical schools: how to succeed." Academic Medicine **75**(6): 575-594.

Boet, S., A.-A. Pigford and V. N. Naik (2016). "Program director and resident perspectives of a competency-based medical education anesthesia residency program in Canada: a needs assessment." Korean J Med Educ **28**(2): 157-168.

Cohen, E. R., W. C. McGaghie, D. B. Wayne, M. Lineberry, R. Yudkowsky and J. H. Barsuk (2015). "Recommendations for Reporting Mastery Education Research in Medicine (ReMERM)." Academic Medicine **90**(11): 1509-1514.

Cook, D. A. (2010). "Twelve tips for evaluating educational programs." Medical Teacher **32**(4): 296-301.

Diamond, R. M. and A. Sudweeks (1980). "A Comprehensive Approach to Course Evaluation." Journal of Instructional Development **4**: 1.

Lewis, C. (2015). "What Is Improvement Science? Do We Need It in Education?" Educational Researcher **44**(1): 54–61.

Lombarts, M. J. M. H. (2015). "Competence based education misses the essence of the medical profession." Perspectives on Medical Education **4**(6): 326-328.

McGaghie, W., J. H. Barsuk and D. B. Wayne (2015). "Mastery Learning With Deliberate Practice in Medical Education." Academic Medicine **90**(11): 1575.

McGaghie, W. C. (2015). "Mastery Learning: It Is Time for Medical Education to Join the 21st Century." Academic Medicine **90**(11): 1438-1441.

Mehrens, W. (1991). Measurement and Evaluation in Education and Psychology (4th Ed). NY, NY, Holt, Rinehart & Winston.

Miller, G. E. (1990). "The assessment of clinical skills/competence/performance." Academic Medicine **65**(9): S63.

Pangaro, L. and O. Ten Cate (2013). "Frameworks for learner assessment in medicine: AMEE Guide No. 78." Medical Teacher **35**(6): e1197-e1210.

Rahim, A. and R. A. Anderson (2008). "Strategies for Residency Programs that Improve Medicine Departments and Teaching Hospitals." The American Journal of Medicine **121**(5): 450-455.

Schultz, K. and J. Griffiths (2016). "Implementing Competency-Based Medical Education in a Post-graduate Family Medicine Residency Training Program: A Step-wise approach, Facilitating Factors, and Process or Steps That Would Have Been Helpful." Academic Medicine **91**(5): 685-689.

The Royal Australian College of Physicians (2015). When learning at work is not enough: Embedding medical education in the DNA of organisational systems and physician practice. Australia, The Royal Australian College of Physicians.

8. Appendix 1: Annotated Bibliography

Vasser, M., D. L. Wheeler, M. Davidson and J. Franklin (2010). "Program Evaluation in Medical Education: An Overview of the Utilization-focused Approach." Journal of Educational Evaluation in the Health Professions 7(1): DOI: 10.3352/jeehp.2010.3357.3351.

Summary

The paper discusses a range of program evaluation models, including an objectives-oriented approach, a management-oriented approach, participant-oriented approaches, expertise-oriented approaches, and logic models. It argues for a utilization-focused approach, as it assumes that evaluations should be judged on their actual use and utility. Utilization-focused approaches are context-specific and flexible, and can therefore incorporate elements from any program evaluation model. They are systematic, starting with identification of the primary users, followed by a commitment from the users and the evaluator to address particular goals, and culminating in the selection of the research design and measurement approach. The last phase should include: consideration of sampling issues, the nature of the data collected, and the quality of the assessment devices. The final step also requires decisions about the dissemination of the evaluation report.

Implications

The utilization-focused approach is highly relevant to the design of assessment tools to monitor achievement and progress in a CBME model. Its flexibility is essential in the context of CBME, where measurements will occur in a very programmatic context that is unique to the program, competency, and setting. A utilization-focused approach includes principal stakeholders in the decision-making process, including learners and faculty, which is critical to developing buy-in for the CBME assessment strategy.

Diamond, R. M. and A. Sudweeks (1980). "A Comprehensive Approach to Course Evaluation." Journal of Instructional Development 4: 1.

Summary

The focus of this article is broad-based, for all disciplines in post-secondary course evaluation. The paper is written in general terms and addresses two specific questions:

1. What purposes and/or roles can be served by formal systematic course evaluations?
2. What are the issues that must be explored, and questions that must be asked, in determining the value of a course?

Synopsis

The article considers advantages and disadvantages of the current system: evaluation is often an afterthought; evaluation raises questions that might otherwise be avoided; evaluation instruments can define the scope and focus of the evaluation process; there is no single approach to evaluation; and a wide range of options must be provided in order be able to select the most suitable one(s). The article concludes with a comprehensive checklist that serves as a functional guide for those faculty and administrators with responsibility for conducting course evaluations.

Implications

Although the paper is general with respect to course evaluation, the checklist is a useful framework to consider as programs start to develop tools both for assessment of learners and for learners to use in evaluating the learning experience. The checklist could be adapted for PGME use. The article brings forward some important concepts around constructing an assessment framework for CBME, such as designing more comprehensive and revealing evaluations, posing appropriate questions, and considering the timing of evaluations and accompanying faculty training.

Cohen, E. R., W. C. McGaghie, D. B. Wayne, M. Lineberry, R. Yudkowsky and J. H. Barsuk (2015). "Recommendations for Reporting Mastery Education Research in Medicine (ReMERM)." Academic Medicine 90(11): 1509-1514.

Summary

The authors reviewed the literature and identified 38 guidelines in 22 categories that can be used to assess studies of mastery learning. Mastery learning is viewed as essential to the implementation of CBME. This paper takes for granted the superiority of a CBME ("mastery") model to the traditional time-based apprenticeship model, as the latter leads to variable acquisition of skills and the propagation of deficient abilities from one generation of learners to the next.

Synopsis

The authors identify seven attributes of mastery learning: 1) baseline testing; 2) clear objectives; 3) activities designed to support achievement of these objectives; 4) minimum passing standards; 5) formative testing; 6) advancement contingent on achieving standards (which may take a variable period of time); and 7) continued practice until mastery is achieved. Generally, there is support for the mastery learning approach, but also a recognized need for more rigorous standards for reporting of outcomes. This is consistent with the Accreditation Council for Graduate Medical Education (ACGME) requirements for outcome-based educational milestones. Accordingly the authors reviewed multiple fields for insight into this challenge. They generated a list of mastery reporting guidelines. There are 38 of these guidelines in 22 categories, divided into the six sections of a manuscript: 1) title and abstract; 2)introduction (including literature review); 3) methods; 4) results; 5) discussion; and 6) other information. Many of the guidelines presented are relevant to all medical education

research reports, and are not commented on further here. The ones that are particularly relevant to mastery learning include the following:

- Defensible minimum performance standards, in a criterion-referenced manner, must be established. The standard-setting process must be documented, and the consequences of the learner's not meeting the standard must be specified.
- Details of the educational intervention must be sufficient to permit replication of it in another setting, particularly the core aspects of the mastery learning process. These details include a description of the baseline or diagnostic assessment, the deliberate learning process, the processes used for feedback, and the *a priori* effect size used to establish practical significance of any observed differences.
- Reporting of results needs to include how many trainees achieved mastery level in both the mastery and the conventional groups, and how many in each group required extra time and resources.

The authors acknowledge several limitations and recognise that other educators would likely include additional attributes of mastery. They also acknowledge that there are significant time and resource implications inherent in a mastery learning framework, and that most research to date in this domain has focused on procedural skills, so there is a strong need for research in other domains, such as communication skills, problem-solving, and other cognitive abilities.

Implications

As the individual programs identify educational strategies and measurements to establish efficacy in the context of CBME, the reports and papers describing these interventions should be critiqued using the kind of framework proposed in this paper.

Furthermore, it is presumed that in many cases rigorous empirical data may not be available to provide these assessment modalities, and so programs will need to develop their own measures. The processes followed in doing so should adhere to the procedures outlined in this paper, *as if* the developers were intending to present the procedures and results for purposes of dissemination. It is of course hoped that introducing CBME at U of T (and elsewhere) will generate much scholarship, but whether this occurs in any one particular case or not, the procedures articulated in the paper by Cohen and colleagues seems to be an appropriate developmental model for assessment instruments, particularly with respect to standard-setting and procedures for providing feedback and supporting mastery learning.

Goldie, J. (2006). "Evaluating educational programmes AMEE Guide No. 29." Medical Teacher 28(3): 210-224.

Summary

This paper is an expert review of evaluation theory, from understanding the political landscape to planning, conducting, interpreting, and finally disseminating results. There

are some very useful definitions of “Evaluation” that program directors and residency program committees may find helpful for understanding the breadth of the term: ‘‘the act of judgment of the worth of …’’. As such, it is an inherently value-laden activity (Collins English Dictionary 1994): ‘‘the systematic process of determining the extent to which instructional objectives are achieved’’ (Gronlund 1976); and “an examination conducted to assist in improving a programme and other programmes having the same general purpose (Cronbach 1980).

Synopsis

The author uses some important discriminants between evaluation and assessment as these terms are often (wrongly) used interchangeably; assessment being primarily concerned with the measurement of student performance, while evaluation is generally understood to refer to the process of obtaining information about a course/program of teaching. Assessment is therefore a subset of evaluation (Newble and Cannon1994).

Mehrens’ (1991) work identifies the purposes of assessment as being:

- to evaluate the teaching methods used;
- to evaluate the effectiveness of the course.

Another broad concept that we can disseminate and use to generate discussion is the purposes of evaluation. These have been described as falling into three general categories:

- evaluation for accountability;
- evaluation for knowledge;
- evaluation for development.

Implications

As a result of the discussion we think that the approaches to a program evaluation are the most useful in the context of initiating CBME for programs. The different approaches emphasize different aspects of evaluation with identified priorities. The broad categories are: objectives-oriented, management-oriented, consumer-oriented, expertise-oriented, adversary-oriented, participant-oriented. There is a useful table describing the limitations for each, as well as a distribution of them on a utilitarianism-intuitionist/pluralist spectrum.