

Medical Education Assessment Advisory Committee

Report to the Medical Council of Canada on Current Issues in
Health Professional and Health Professional Trainee Assessment

Kevin Eva (Chair)	University of British Columbia
Georges Bordage	University of Illinois at Chicago
Craig Campbell	Royal College of Physicians and Surgeons of Canada
Robert Galbraith	National Board of Medical Examiners
Shiphra Ginsburg	University of Toronto
Eric Holmboe	American Board of Internal Medicine
Glenn Regehr	University of British Columbia

April 25, 2013

Foreword

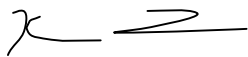
On behalf of the Medical Education Assessment Advisory Committee (MEAAC), I am pleased to submit this report entitled Current Issues in Health Professional and Health Professional Trainee Assessment. As a new committee serving the MCC directorate, being charged with this task offered a useful stimulus to begin our conversations about ways in which the assessment practices of the MCC and other related organizations establish a culture (for better and worse) of what it means to practice medicine in Canada.

While our discussions to date have been far reaching and have not been constrained to MCC practices in particular, none of the contents contained in this report should be read as critical of the MCC. On the contrary, the very facts that MEAAC exists and that the MCC is engaged in a genuine and extensive exploration of ways in which it might evolve and or revolutionize its assessment practices we see as very positive indicators that the MCC is dedicated to fulfilling its mandate to the best of all possible standards.

We are grateful to Ian Bowmer, Claire Touchie, Andre de Champlain, and many other members of the MCC staff for helping to educate us about the MCC's current practices and the many projects being undertaken for the continued development of a strong system of assessment; a system that should certainly "protect Canadians while sustaining public confidence in our physicians" as noted on www.mcc.ca to be the essence of the MCC's story.

Having recently celebrated its 100th birthday, we think the MCC to be well positioned to continue to thrive over the course of its next centenary and are proud to have had the opportunity to play a small role in helping to set its course for the near and intermediate future. We see this report as the start of our conversation with the MCC rather than the end and look forward to this continued engagement.

Sincerely,



Kevin W. Eva
Chair
Medical Education Assessment Advisory Committee

Table of Contents

Executive Summary	3
Background and Rationale	4
Current issues facing health professional testing organizations.....	5
Theme 1:	
Overcoming unintended consequences of competency-based assessment	7
Logistical Considerations	8
Theme 2:	
Striving to implement quality assurance efforts in a manner that facilitates learning and performance improvement.....	10
Logistical Considerations	10
Theme 3:	
Authenticity in assessment practices	13
Logistical Considerations	14
Systemic Considerations	16
Summary of Recommendations	17
Appendices	20
Context	20
MCC Blueprinting Project.....	20
Medical Education Advisory Assessment Committee.....	21
Our task.....	21
Our approach	21

Executive Summary

The fundamental purpose of this report is to offer insights into ways in which the Medical Council of Canada might adapt its current assessment practices and develop additional practices to ensure alignment with modern conceptions of medical education and medical practice, the ultimate goal being better healthcare for Canadians.

To do so, the Medical Education Assessment Advisory Committee engaged in a series of iterative discussions that could be condensed into three dominant themes:

- 1) The need to redress unintended consequences of competency-based assessment;
- 2) The potential to design assessment systems that facilitate performance improvement; and,
- 3) The importance of ensuring authenticity in assessment practices.

The following principles were found to cut across each of these themes and represent the foundational goals we would put forward as signposts for decision-making about the continued evolution of the MCC's assessment practices:

- 1) Creating greater opportunity to enable assessment practices that promote learning rather than simply measuring performance;
- 2) Enabling integration across stages of training/practice and testing organization/jurisdiction to smooth out the transitions experienced by individual trainees/practitioners; and
- 3) Moving away from exclusive reliance on point-in-time assessments to prioritize continuous professional development in a manner that enhances a shared model of responsibility/accountability between practitioners and educational programs/testing organizations.

At this point in time many of the ideas generated through this exercise represent concrete suggestions for strategies to pilot test and for infrastructure to build rather than explicit recommendations that can or should be implemented immediately. These include novel strategies for OSCE station development, formative and “diagnostic” testing protocols tailored to shed light on the practices of individual clinicians, the use of continuous workplace-based assessment, and broadening the focus of decision-making beyond determining who passes and who fails the various examinations mounted by the MCC. The report concludes with reflections on systemic barriers that may need to be overcome to move towards a more integrated, efficient, and effective system of assessment as well as a series of specific recommendations for steps to be taken to facilitate movement in that direction.

Background and Rationale

The context that led to the generation of this report, the methodology used, and the function of MEAAC is outlined in an appendix. The foundational principle on which this report was built is that adequate consideration of the blueprint of any assessment system requires more than simply generating a process and content rubric to which administrators should adhere. That is, not all issues of relevance to the assessment of health professionals boil down to the simple question of what is assessed or how. In addition, it is important when deciding upon adoption of an assessment system to consider both the implicit messages sent to candidates and stakeholders and any unintended consequences of adopting particular assessment strategies. This is Messick's notion of consequential validity and requires a broader, scoping, review of not only current best practices in assessment, but also the alignment between current assessment systems and modern understanding of health professional practice. It is this aspect of the MCC Blueprint Project that the MCC's Medical Education Assessment Advisory Committee was asked to address and which forms the basis of this report.

Our goal was not to debate the merits of any particular form of assessment. Nor did we allow ourselves to be constrained by the particulars of the current structure of the medical licensing process used in Canada. This report is not based on an empirical study or a systematic review of the literature. Rather, it is intended to represent a consensus-based statement of current issues that came to light by virtue of the connoisseurship of a diverse group of individuals who have dedicated considerable portions of their variable professional lives to understanding the intricacies of assessment practices within the health professions. Across its members the group is well versed in (and experienced with) the day-to-day assessment practices of medical schools, licensing authorities, certifying bodies, and regulatory offices in North America and further abroad. They have held many senior positions, are well-respected scholars in health professional education, and have contributed extensively to knowledge growth efforts in this domain.

The general themes this process raised, three of which are highlighted as the basis of this report, do not necessarily reflect issues that have simple solutions. Where possible we have tried to offer examples of how broader conceptual issues could be experimented with within the MCC assessment context, but specific suggestions should be read as an effort to offer concrete paths for exploration rather than absolute guidelines regarding how assessment processes should be implemented. Each theme reflects on the culture of medical education and medical practice in Canada at the time this report was generated.

We think the MCC to be uniquely positioned to bring together key stakeholders to drive the continued development of health professional education assessment practices and support the Assessment Review Task Force's recommendation that the MCC prompt further collaboration between itself and other national physician assessment bodies. With the imminent roll-out of the harmonization project with the College of Family Physicians of Canada, the development of the National Assessment Collaboration, and the ongoing procedural review processes being engaged by many related organizations, the timing appears to be ripe for further developments of this type. Before addressing the specific themes arising from MEAAC's discussions, therefore, it is valuable to consider both current trends and key issues that are common to all testing organizations operating within the domain of health professional education assessment.

Current issues facing health professional testing organizations

With assessment practices and conceptions of health professional practice and education evolving markedly over recent decades it is important to specify the context within which current issues in health professional assessment are being evaluated. We consider van der Vleuten's model of utility (reliability, validity, feasibility, acceptability, and educational impact) to still provide a useful model from which to judge the adequacy of any assessment system and would emphasize, as he did, that compromise is necessary across these factors. In an ideal world that compromise will result from deliberate decision-making rather than uninspected imbalance. We recognize that the MCC offers a high stakes testing environment that must prioritize reliability (consistently discriminating between candidates, particularly at the pass-fail decision point) and validity (ensuring that the data are fit for the purpose of drawing the inferences that the assessment body desires to make) for the sake of both fairness to the candidates and the safety of patients. It is for this reason that multiple choice question (MCQ) exams of medical knowledge have become the mainstay of licensing examinations in North America and are likely to remain a fundamental component of health professional assessment practices. At the same time, the broadened conception of validity that Kane, Messick, and others have put forward demands that a wide variety of additional factors be taken into account when deciding whether or not a system of assessment is fit for purpose.

Miller's pyramid (and the related explorations that have arisen since its publication) along with efforts to develop comprehensive frameworks of medical practice (such as those presented in CanMEDS) continue to emphasize the value of expanding the assessment of health professionals beyond simply testing medical knowledge. The MCC has a strong history in this regard given its incorporation of Considerations of Ethical, Legal, and Organizational (CLEO) aspects of practice into its examinations and its use of Objective Structured Clinical Examinations to capture candidates' ability to demonstrate important and broad clinical behaviours such as communication skills. We see these innovations as steps in a process that should continue. The MCC is not alone in this regard as each and every testing organization within the health professions should continue the effort to broaden the base of assessment beyond knowledge tests.

Advocacy of such efforts should not be construed as simply indicating that additional aspects of practice need to be measured or that additional types of assessment need to be used as part of the MCC's examinations. Rather, we see such efforts as demanding a more coherent and integrated system of assessment across the continuum of training and practice such that the process (a) is made efficient for candidates, (b) emphasizes the primacy of learning by harnessing the power of feedback, (c) creates a shared responsibility/accountability between the learner and the educational program/assessment system by ensuring that information about one's strengths and weaknesses travel with candidates from one stage of their career to another in a manner that enables and promotes continuous performance improvement, and (d) ensures appropriate and comprehensive coverage of performance while eliminating unnecessary redundancy of effort across the assessment practices of medical schools, licensing/certification bodies, and regulatory authorities.

It likely goes without saying that all of these efforts need to take place within the context of a rigorous quality assurance effort, but we would encourage that 'quality' in relation to many of these issues be considered in a manner that takes into account more than the most common

psychometric markers of test performance. The integration efforts described above should enable databases to be generated that allow added emphasis to be placed on outcomes of clinical care (and the capacity of the assessment system to predict those outcomes). At the same time, they could also be used to enable encouragement towards (and examination of) meaningful engagement in continuing professional development and successful interaction with other members of the healthcare network.

In sum, we consider the following to be a list of common issues facing high stakes testing organizations in the health professional community:

- (1) Broadening the base of assessment beyond knowledge tests
- (2) Rigorously focusing data collection and decision-making practices in a manner that enables the assessment body to draw relevant and meaningful inferences
- (3) Adding emphasis on healthcare processes and outcomes, including strengthening of the ability of the assessments to predict who will perform well against those outcomes and who will further develop in their ability after training
- (4) Building a coherent and integrated system of assessment across the continuum of training/practice
- (5) Emphasizing the primacy of learning as an integral part of assessment
- (6) Harnessing the power of feedback
- (7) Shifting accountability towards a model of shared responsibility between the individual and the education program/system.

In an effort to balance the discussion of the themes included in this report between broad perspectives on assessment and the operational needs of the MCC we have focused on three levels of consideration: (1) Philosophical – big picture conceptual issues about how, why, and when different assessment practices impact upon the culture of the profession and unintended consequences that might arise; (2) Logistical – specific avenues of exploration through which these conceptual issues might be redressed within the practical realities of needing to mount a licensing process; and, (3) Systemic – issues inherent in the current Canadian medical and medical education system that set barriers that may need to be overcome if one hopes to advance assessment and educational practice beyond their current state.

Theme 1: Overcoming unintended consequences of competency-based assessment.

Modern systems of medical education are increasingly centred on competency-based frameworks that outline the core knowledge and skills that individual physicians are expected to maintain. This movement has had definite positive impact by offering an explicit and broadly focused model from which educators, regulatory authorities, and licensing bodies can guide educational and assessment practices for the sake of improved patient care. In doing so it has increased the emphasis on trainee and physician observation given that many of the aspects of practice listed in common competency frameworks are not easily amenable to measurement in other ways. While there are well-recognized psychometric challenges for many of the assessment strategies that rely on such observations, in general these challenges can be overcome through deliberate and diverse sampling of the competencies across different situations. Also, the observations offer substantial potential for pedagogical benefits that range beyond sound measurement including increased opportunity for formative feedback.

Implicit in common models of competency-based assessment, however, are a variety of assumptions that may have unintended and undesirable consequences for medical education and medical practice. Most central is the notion that competence is something one can achieve, full stop. Claiming that a student can “perform a complete and appropriate assessment of a patient,” for example, ignores the robust literature indicating that contextual factors play an important role in our ability to perform any task and sends an implicit message that once a task can be achieved there is no further work to be done. As mentioned above, the issue of contextually driven variability in performance can be overcome by sampling performance broadly. The latter issue of the implicit messages sent by adopting a competence-oriented assessment strategy (and the MCC’s exclusive focus on minimal competence thresholds) has a more fundamental impact on the culture of practice and education. This impact is observable in many ways.

The fact that every candidate who passes an MCC exam is effectively labeled competent for the next stage of training/practice overlooks the realities that (a) there is considerable variability of performance within the passing range, (b) even the top performers have room for improvement, and (c) knowledge and skill are subject to drift and deterioration (decay) over time. A considerable amount of information is collected on students during their time within medical school, postgraduate training, and while undertaking the licensing/certification examinations in Canada. For the most part, this wealth of data is simply ignored when the candidate is, at a particular point in time, deemed ‘good enough’. This creates any number of problems. First, it creates a situation in which examinations and other assessment protocols are seen as hurdles that one simply needs to get over rather than as diagnostic opportunities that can be put to use for further pedagogic benefit. Second, focusing exclusively on the cut-point at which it is determined who passes and who fails discourages trainees from disclosing difficulties they might be having as there is no incentive (and considerable disincentive) for continuing to pursue improvement of weaknesses as long as one has successfully made it over the hurdles they face. In fact, passing the examination is likely to be treated as an implicit indicator that the weaknesses one experiences are relatively unimportant given that passing means the profession has deemed the learner sufficiently knowledgeable. Third, such models may simultaneously reduce the degree of support that educators feel compelled to provide given that there is little need to observe or offer guidance to trainees who have been deemed competent (or who have at least

been deemed to be on the path toward competence). Finally, reinforcing the perception that one can be labeled as ‘competent’ at any given point in time overlooks the well-established view that knowledge and skills must be continuously reactivated (i.e., used) for them to be maintained. Having successfully crammed to pass an exam should not be viewed as an indication of the likelihood that one will have the knowledge or skills that were tested available in actual practice well after the exam was completed.

Each of these problems creates risk in that the overarching goal of individuals within such an environment is to pass the examination (i.e., to “achieve a state of independence”), while the degree to which actual practice is truly independent is lessening and the emphasis on continuous efforts at performance improvement is increasing. That is, this “state of independence” runs counter to modern perspectives on expertise, which differentiate between the routine expert who achieves a certain degree of performance and simply reproduces that performance repeatedly and the adaptive expert who continuously reinvests any mental resources freed up by the automatization that accompanies experience into better understanding their domain of practice and innovating for the sake of continuous performance improvement. It has been demonstrated in countless domains that deliberate practice is necessary to reproducibly improve one’s skill level beyond its current state and that reduction of such effort results in performance declining back to previous ability levels. This is a greater risk in medicine than it is in most domains as the knowledge/skill base on which medical practice is built is constantly changing.

All of this sums together to create a state in which focusing efforts exclusively on the few individuals who do not meet the threshold eliminates opportunities to provide formative guidance directing future learning for the approximately 97% of (Canadian trained) graduates who are able to pass the licensing examination on their first attempt. While the 3% definitely need to be excluded from practice until such time as their ability can be improved, formally supporting the 97% in their performance improvement efforts could arguably have a bigger impact on the quality of healthcare that patients receive in Canada.

Logistical considerations

While emphasis on direct observation and accompanying feedback in medical school, postgraduate training, and even continuing professional development is increasing, this must be considered within the reality that most physicians are well intentioned yet overworked. It is not uncommon, as a result, for attending physicians to convey conflicting messages to their trainees by, for example, stating “call me if you have any problems” while simultaneously disclosing feelings of exhaustion that accompany long periods of work. Rarely, in fact, are trainees observed, especially when they are on call or seeing new patients for the first time. It is not reasonable, as a result, to expect that trainees would be observed constantly. Nor would that be ideal given the value of progressive independence and of desirable difficulties (being challenged in a manner that drives learning) for performance improvement. A careful delineation of Entrustable Professional Activities (EPAs) at various stages of practice may help facilitate identification (and more consistent awareness) of activities that do require relatively less observation while simultaneously offering some guidance regarding mechanisms through which multiple competencies might be integrated.

That said, what we are primarily advocating for by drawing attention to the unintended consequences of competency-based and individually-focused assessment is that we as a field can potentially do better at optimizing the time and resources that are available. A more integrated and continuous model of assessment across the various stages of training/practice in which information is carried forward with the individual would optimize the profession's capacity to hone in on the particulars of performance that would be most impactful for a particular learner at a particular time and promote the notion that trainees need to be accountable and maintain 'ownership' of their own learning. An assessment system that recognizes the continuous nature of performance (as opposed to dichotomizing into pass-fail) would further normalize this process such that all learners (regardless of whether they "pass" the previous assessments) would maintain a learning plan that could efficiently guide both their own and their preceptors' activities in a manner that could minimize wasted time and energy. Thinking this way about competent practice will not remove the need to develop strategies to enable further opportunities for observation of supervisors by trainees and vice versa, but it may create a situation in which such opportunities are more rewarding for all parties as they can be more deliberately directed at the key aspects of a trainee's performance that would most benefit from improvement.

To some extent these ideas are translations of the key features model of exam question development used by the MCC in that they lead to a proposal to use the information available in any form of assessment to define the key actions or diagnoses that are most appropriate at a particular moment in one's development. The high quality of the summative assessment processes mounted by (or supported by) the MCC are likely to always be more trustworthy in this regard compared to locally generated assessment protocols that are often implemented without adequate resources in terms of assessment expertise. With respect to current MCC practices, it is conceivable that OSCE stations or key features questions used in the Clinical Decision Making (CDM) component of the MCC's examinations could be built such that they require candidates to follow up on an error that was made; to ask for help, guidance, or supervision; to use a clinical decision support; or to summarize a case for an attending, making it clear not only what is understood, but what the learner would take away from the situation to direct further formative development. Further, it is conceivable that the detailed information collected during the assessment process could be fed back to candidates in a way that makes their intra-individual relative strengths and weaknesses more apparent (or that additional information like confidence ratings might be collected to highlight areas of inappropriate confidence or lack thereof), thus providing concrete guidance regarding how investment into continued learning can be optimally spent. Offering tailored components to subsequent assessment processes that require the candidate to demonstrate how they have utilized such information to their own and their patients' benefit would create considerable incentive for candidates to treat their learning as a lifelong effort requiring continuous reinvestment rather than simply trusting that they know enough because their exams have been passed. In this manner, it is plausible that the goal of building a coherent and integrated assessment system could be achieved in a way that addresses the further goals of shifting accountability toward a shared responsibility between learner and system while also harnessing feedback and emphasizing the primacy of learning. This emphasis on performance improvement rather than simply focusing on quality control is the focus of the next theme that arose from MEAAC's discussions.

Theme 2: Striving to implement quality assurance efforts in a manner that promotes learning and performance improvement

For assessment of any kind to provide a meaningful mechanism through which individuals can be expected to grow it needs to promote, support, and empower self-regulating aspects of practice, provide high quality and credible feedback, and deliver coaching/support for the candidate. We understand the argument that high stakes testing organizations are responsible for measurement of performance and cannot be expected to prioritize assessment for learning. We do not, however, believe it. Part of the challenge we see with current assessment practices, as alluded to earlier, is that they are singular events with no effort to facilitate use of the information gained from their administration and no effort directed at follow-up to determine if the information gained has been utilized. The very fact that multiple roles inherent in modern practice (e.g., Scholar, Professional) incorporate competencies that require individuals to continually reinvest in developing their practice over the course of their careers combined with the robust evidence that introspection is misleading in this regard demands that good data be sought by the individual practitioner regarding his/her current level of practice.

Given that the expertise in assessment and the resources available for assessment practices are substantially greater in testing organizations like the MCC than they are in most other institutions that physicians will encounter in their medical career, it will be rare for trainees or physicians in practice to ever have better data on which to guide their continued development than they do upon engaging in formal assessment activities. To not consider the use of assessment for performance improvement in this context is, therefore, a considerable missed opportunity. No effort is currently made to assess (or even promote) feedback seeking behaviours or responsiveness to feedback, yet those capacities may be the most important factors in determining continuing competence during graduate medical education (GME) and accompanying practice improvement efforts. There will inevitably be a tension between quality control and quality improvement goals whenever the threat of high stakes assessment looms over candidates. With further integration across the continuum of training, however, and with deliberate attention paid to quality improvement as a goal, we believe benefit can occur despite such tension. Thus, we offer some initial thoughts regarding how such concepts might be built into the activities of the MCC.

Logistical considerations

In contemplating the issue of the MCC supporting performance improvement, MEAAC generated a number of possible activities, some of which are described here. They are presented in an order that corresponds roughly with the chronological sequence inherent in the continuum of training and practice. It is not thought that the MCC could (or even should) mount each of these activities, but it is thought that the expertise inherent in the MCC could be put to further use in supporting individual medical schools and collaborating with other institutions to create an integrated system of assessment that yields genuine guidance to enable performance improvement.

A burgeoning area of research in recent years is demonstrating the pedagogical value of testing. A number of studies have demonstrated that material that has been tested is often more

memorable than the same material after it is simply studied multiple times. As with all forms of memory retrieval (as alluded to earlier in reference to the importance of reactivating knowledge and skill), more testing tends to yield greater effect, especially when the testing format requires constructed responses (e.g., short answers) compared to recognition (e.g., MCQs). This phenomenon creates a perspective in which formative quizzes and longitudinal (progress) testing models become increasingly valuable. The cost of generating test items for the licensing examination makes support of these activities prohibitive using traditional practice, but the MCC's investment in automatic item generation (AIG) processes should make progress testing much more feasible without the need to worry as extensively about maintaining the security of items used for that purpose.

Once AIG techniques are enabled to generate a database of questions, we would advocate the creation of a test tailoring platform in which individual students and physicians could specify the domains on which they are currently focused and draw a test for themselves in live time that would both support their own learning and provide them with habits and evidence of engaging in deliberate practice improvement activities that could be carried on throughout the course of their careers. In an ideal world, at least at the GME level, databases would be created that would allow practitioners to sync their current practice profiles (based on electronic health records, prescription habits, etc.) to a rubric that defines the AIG database such that formative tests could be drawn that are optimally tailored to each individual's practice and would yield guidance regarding mechanisms of improvement. The success of such guidance could then be tracked through the individual's practice patterns by considering the very patient outcomes that led to the generation of the formative test in the first place.

Enabling and encouraging this type of activity can be facilitated by generating a licensing examination protocol in which one component requires candidates to demonstrate how feedback on areas requiring improvement were identified prior to their examination (or, in subsequent assessments, as a result of their examination) and led to an action plan that has formed the basis for their learning (and will inform the continuation of their learning once in practice). An OSCE station, for example, that involves review of data/information collected from each candidate's actual patient encounters (and requires them to demonstrate how they have aggregated their experiences and used data built up prior to the licensing examination through workplace-based assessments, evaluation of patient outcomes, and summary of practice experiences) would seem possible while also extending the scope of the licensing examination away from the single moment in time in which the candidate is physically present for the exam.

Continuing this line of thought, the MCC might be well positioned to support a "Diagnostic OSCE" late in undergraduate MD training or early in postgraduate training, blueprinted on key aspects of practice (perhaps Entrustable Professional Activities), that could be used to identify aspects of performance that (a) would benefit from further development and (b) could form the basis for tailoring subsequent assessment efforts aimed at exploring how relative weaknesses have been strengthened. An exercise like this might be particularly valuable late in MD training if schools move towards having their students sit Part I of the Qualifying Examination earlier in the year (as the MCC addresses the ARTF recommendation of increased flexibility with respect to the timing of the exams). It would arm both students and postgraduate training programs with concrete guidance regarding activities that might be prioritized early in residency as a means towards efficiently helping trainees progress rather than waiting for relative weaknesses to be re-

discovered throughout the postgraduate training process. Similar benefits could be achieved by having international medical graduates (IMGs) complete the same diagnostic OSCE as part of the National Assessment Collaboration, thus giving postgraduate training programs more information with which to help IMGs adjust to the Canadian healthcare system. The AIG protocols mentioned above could be used to populate databases of stations for this purpose for use by individual medical schools and the resulting data could enter into a candidate controlled portfolio with a requirement to report out to the MCC and/or certification bodies for the sake of tailoring follow-up stations/assessments as part of later stages of the licensing/certification/maintenance of certification processes.

Theme 3: Authenticity in assessment practices – reflecting the full domain of practice

Included in MEAAC's view of the utility of the MCC's assessment protocols is that the system of assessment should model the realities of practice as closely as possible. Such alignment increases acceptability and makes claims of validity much more credible. We do not, however, think that authenticity necessarily means using high fidelity simulation to mimic actual practice. Rather, by emphasizing the value of authenticity we mean to say that it is important that assessment represent practice for the sake of ensuring that the gamesmanship that is inevitable with any high stakes assessment process results in actual performance improvement. That is, if an assessment protocol accurately reflects the domain of practice, then "studying to the test" or learning to "game the system" should equate with learning to practice well. Too often we hear statements from clinical preceptors to their trainees along the lines of "in reality I would do X, but for your exam you should do Y." Such disconnects threaten to undermine the entire system. They reinforce the concerns expressed earlier that the exams are treated merely as hurdles to be overcome as a way of proving oneself competent, they encourage trainees towards practicing in a manner that does not align with reality, and they encourage educational programs to set curricula/teaching practices to the exam rather than to the professional careers of its students.

Given the well-recognized and well-grounded adage that assessment drives learning, this aspect of MEAAC's deliberations may be the most fundamental to enabling the cultural shift within the profession that has been strived for through modern efforts to broaden the definition of good medical practice. In fact, we see this as being so central to fulfilling the goals of the MCC's blueprinting effort that we think surveying upcoming candidates and current faculty/residents about what they predict will be included on the next licensing examination (i.e., what they would advise a test candidate to expect) could create a baseline from which the impact of future test development efforts could be measured. Whether responses focus on superficial case types or deeper conceptual issues, the variety of competencies reflected in participants' responses, and the alignment between responses and the MCC's actual objectives would all provide guidance regarding the extent to which the MCC's assessment practices and its advertising of those practices are effectively aligned with modern conceptions of authentic medical practice. "How predictable have the exams become" is a validity question in its own right as relevant stakeholders should be able to demonstrate an accurate awareness of the goals of the licensing process just as the process should accurately reflect the aspects of practice the MCC and its stakeholders desire to promote.

Workplace-based assessment practices including in-training evaluation reports, mini-CEX, patient reported outcomes, and direct observation of procedural skills, in many ways the final frontier of assessment technology, are not currently part of the MCC's assessment activities. We think it important that they be considered given their potential for assessing a greater variety of dimensions of practice and their capacity to better reflect what individuals actually do in their day-to-day activity. We appreciate that there remain concerns about standardization and reliability with most workplace-based assessment practices, but again, think this to be one domain in which the benefits may reside in aspects of assessment practice that range beyond achievement of a high reliability coefficient. In the subsequent section we try to illustrate these

potential advantages through concrete thoughts regarding how these practices might be implemented to good effect.

Logistical considerations

At its root, making assessment authentic requires having candidates engage with clinical scenarios that are not clear and obvious cut-outs from a blueprint. That is, the cases must allow uncertainty and avoid prompting statements such as “here comes the breaking bad news station.” Doing so might involve allowing stations with multiple pathways even at the cost of absolute standardization. Possible models are being developed as many groups around the country have been experimenting with sequential OSCE processes in which later stations revisit previously encountered patients at an ostensibly later point in time, offering test results that were not available previously or some other form of follow-up visit. Within station, it is also conceivable that standardized patients could be trained to offer information midway through a case that contradicts the most apparent diagnosis/key feature from the early portion of the encounter, as has been done with Clinical Decision Making cases in the past. Doing so would further provide some indication of candidates’ capacity to overcome their first impressions and avoid falling prey to premature closure.

At the same time, when using rater-based assessment strategies such as the OSCE or Mini-CEX, we tend to infer the cause of certain behaviours, trusting that the right things, when done, were done for the right reasons. Given that it is well established that context influences performance there may be value in establishing opportunities for examiners to explore the reasoning underlying candidates’ behaviour. This could be done through post-encounter probes that require the candidate to explain why certain things were done, why alternative actions were ruled out, and if or how his/her decision-making might have changed if the context had differed in specified ways. Gaining a better understanding of candidates’ conceptualizations of practice may help to account for some of the apparent inconsistency in performance from one station to the next (i.e., context specificity) while also enabling a strategy for assessing other aspects of competence embedded within the Scholar and Professional roles.

At another level entirely, the increasing prevalence and use of electronic resources leads us to believe that inclusion of an Internet enabled computer in all OSCE rooms might better represent current practice by enabling candidates to look things up and model their capacities to search efficiently for relevant information and to communicate effectively with their patient regarding what they are looking for and why. This is an increasingly important issue as physicians now need to routinely demonstrate the ability to interact with patients and their Electronic Medical Records simultaneously. More generally, allowing candidates to demonstrate how to effectively seek external consultation from any source may better position the examiner to consider whether or not the candidate acts in a way that maximizes patient safety while simultaneously reducing the implicit message outlined in the first theme that practicing medicine effectively in modern times requires doing so independently. Other testing organizations are known to be experimenting with presenting mock pharmaceutical advertisements to candidates to enable an assessment of their ability to critically appraise information and interact professionally with patients while doing so. Whether or not any one of the potential suggestions listed in this section of the report proves effective in adequately measuring candidate performance, movement beyond simple “examine the knee” types of cases is thought necessary to enable greater potential to

examine holistic aspects of practice that simultaneously take into account multiple competencies (e.g., communication, collaboration, professionalism) and avoids the atomization of medical practice that the specification of a series of competencies can create.

Systemic considerations

A common criticism of the medical training system is the sharp transitions experienced when moving from pre-clerkship to clerkship, from clerkship to postgraduate training, and from postgraduate training to practice. Some degree of transition pain is inevitable, but the challenges created might be reduced by efforts to create a cohesive system of assessment across stages of training/practice. Enabling supervisors, mentors, program directors, and colleges to receive high quality information regarding each individual's relative strengths and weaknesses such that further educational opportunities might be crafted efficiently, would maximize the opportunity to resolve any issues before the next stage of training is experienced and the process of discovery begins anew.

We see value in leaving control of learner portfolios in the hands of the learner to engender a sense of accountability and responsibility. Encouraging active engagement with such activities, however, will require a reward structure to be built that allows data and candidate responses to those data to be recognized as evidence that continuing professional development is being undertaken. At the same time, incorporating review of portfolio-based information and any performance improvement experiences/plans into licensing protocols, peer review efforts, and training programs may better enable program directors and licensing/certification bodies to monitor performance improvement efforts and intervene as necessary before larger problems arise. Establishing a system of this sort will require considerable collaboration between various organizations, but such collaboration appears to be valued by many of the key players and is more achievable in Canada than it would be in many larger countries. Such effort could go a long way towards establishing assessment protocols that authentically reflect the practice in which physicians are engaged.

Any system that relies on continuous or longitudinal forms of assessment and aggregation across many experiences will require better infrastructure than currently exists if one is to truly consider the design of a properly integrated system. This is a challenge that will be difficult to overcome given that it will be complicated by there being varying systems and jurisdictions in Canada. Necessary enabling structures should include an e-learning portfolio with private, public, and de-identified layers, the latter being used for aggregation across physicians for program evaluation and system improvement purposes. Such a structure can help to promote other aspects of an assessment system designed around the issues discussed in this document by forcing integration from matriculation to retirement and requiring collaboration between testing organizations. We recommend that the MCC work with one or two other key organizations such as the Royal College of Physicians and Surgeons of Canada and/or the College of Family Physicians of Canada (CFPC) to pilot test strategies for resolving the complexities inherent in such collaborations. The harmonization project between the MCC and CFPC is good proof of concept that such collaboration is possible.

Summary of recommendations

As alluded to throughout this report, continuing the evolution of Canadian health professional assessment practices in a manner that enables integration across stages of training, emphasizes learning and performance improvement, and creates a more continuous strategy that continues to broaden the scope of assessment beyond medical knowledge while simultaneously sharing responsibility for data-informed continuous professional development will require time, energy, and resources. At the same time, patient safety challenges and the licensing/certification of physicians is not going to come to a stop while these issues are being resolved. As such, we would advocate for an approach that involves engaging a set of pilot efforts aimed at quickly determining the feasibility of multiple strategies that might facilitate movement in the desired direction rather than waiting for a comprehensive system of assessment to be designed and investing heavily in a complete infrastructure, the components of which will undoubtedly be variable in effectiveness. Suggestions for such pilot studies are offered throughout this report. Here they are brought together into a single list organized by the principles outlined in the Executive Summary for ease of retrieval.

Creating greater opportunity to enable assessment practices that promote performance improvement rather than simply measuring performance

1. Develop a strategy for generating a tailored learning plan for individual candidates based on diagnostic information regarding strengths and weaknesses that might be extractable from existing licensing processes.
Purpose: Offering greater guidance to candidates regarding areas for improvement and normalizing the treatment of the licensing process as being part of a continuous process of professional development. This does not mean simply feeding candidates test subscores, but may require a relative ranking strategy and/or establishing a ‘conceptual rubric’ for exam questions distinct from the objectives blueprint used by the MCC.
2. Continue the development of Automatic Item Generation (AIG) processes.
Purpose: Enabling a feasible system for progress testing within medical schools and self-directed assessment practices for individual learners (trainees and physicians) with the ultimate goal of creating a platform through which individuals can tailor a test to their own practice/educational focus in support their own learning.
3. Design a low stakes “Diagnostic OSCE” for use late in undergraduate MD training or early in postgraduate training, blueprinted on key aspects of practice (perhaps Entrustable Professional Activities).
Purpose: Facilitating trustworthy identification of learning needs early in one’s transition to postgraduate training with the potential goals of (a) providing candidates with information they can carry into residency to guide further educational efforts and (b) using the needs identified to tailor subsequent assessment efforts aimed at exploring how individuals’ relative weaknesses have been strengthened.
4. Design OSCE stations with multiple pathways, either through sequential OSCE processes or multiple “decision points” within a station.
Purpose: Exploring candidates’ capacity to think through clinical encounters (rather than simply enabling success based on recognition of station type), the influence of contextual variables, and their personal conceptions of practice.

5. Design OSCE stations in which Standardized Patients are trained to subtly offer information midway through the case that contradicts the most apparent diagnosis/key feature from the early portion of the encounter.
Purpose: Providing indication of candidates' capacity to overcome their first impressions, avoid falling prey to premature closure, recover from mistakes and/or seek out external information in particularly difficult cases.
6. Design post-encounter probes for OSCE stations that require candidates to explain why certain things were done, why alternative actions were ruled out, and if or how his/her decision-making might have changed if the context had differed in specified ways.
Purpose: Gaining a better understanding of candidates' conceptualizations of practice in a manner that may help to reduce measurement error derived from inaccuracy of inferences about candidates' reasoning.
7. Explore the inclusion of an Internet enabled computer/iPad in OSCE stations that will enable candidates to access external resources.
Purpose: Creating an opportunity to observe/measure candidates' capacities to search *in vivo* (live) for relevant information and to communicate effectively with their patient regarding what they are looking for and why. More generally, allowing candidates to demonstrate seeking external consultation may better position the examiner to consider whether or not the candidate acts in a way that maximizes patient safety, thereby reducing the implication that practicing medicine effectively in modern times requires being independent.
8. Survey upcoming candidates and current faculty/residents regarding their perceptions of what will be on the next round of qualifying examinations.
Purpose: Determining the extent to which stakeholder expectations and examining committee intentions are misaligned, thereby providing guidance regarding what issues need to be addressed to improve the authenticity of the MCC's exams.

Enabling integration across stages of training/practice and testing organization/jurisdiction to smooth out the transitions experienced by individual trainees/practitioners

1. Generate a comprehensive map of assessment practices in Canada that takes into account licensing, certification, and the quality assurance activities operated by medical regulatory authorities.
Purpose: Identifying redundancies, contradictions of effort, and opportunities for discussion about collaboratively and deliberately designing a nationwide system that would be integrated across the training-practice continuum and ensure comparability of processes for different cohorts of candidates (e.g., Canadian vs International Medical Graduates).
2. Coordinate a review of e-Portfolio usage within undergraduate and postgraduate training programs in Canada.
Purpose: Determining the common and unique characteristics (along with insights gained) with a view towards judging the feasibility of generating a nationwide system that would reside in the control of the individual candidate (with private, public, and de-identified layers of information) and would enable an individual to both carry learning needs forward in time (and across geographic location/training program) and report out on quality improvement activities undertaken.

3. Identify one or two other key organizations (e.g., certification bodies) to pilot test strategies for resolving the complexities inherent in multi-institutional collaborations.
Purpose: Furthering discussion about ways in which assessment efforts might be merged and/or information from one testing body might feed into the other (likely through a candidate controlled ePortfolio) to enable continuous assessment that requires the candidate to demonstrate engagement with personal performance improvement efforts.

Movement away from exclusive reliance on point-in-time assessments to prioritize continuous professional development in a manner that enhances a shared model of responsibility/accountability between practitioners and educational programs/testing organizations

1. Explore the design of OSCE stations or CDM questions that require candidates to follow up on an error that was made, to ask for help, guidance, or supervision, or to summarize a case for an attending.
Purpose: Such stations or cases could help candidates generate a learning plan from the point-in-time assessments such that follow-up on the experience becomes an important aspect of the licensing process.
2. Explore the development of OSCE stations that require the candidate to demonstrate how they have utilized personally generated information to their own and their patients' benefit. An OSCE station, for example, that involves review of data/information collected from the candidate's actual patient encounters/ePortfolio/peer review.
Purpose: Requiring candidates to incorporate data collection from workplace-based assessments/evaluation of patient outcomes and other quality improvement efforts into the licensing examination to extend its scope beyond the single moment in time in which the candidate is physically present for the exam.
3. Establish strategies for individuals to receive CME credits by engaging with the MCC to develop a low stakes ePortfolio system that would involve data collection and tailoring of learning plans as evidence that continuing professional development is being undertaken.
Purpose: Enabling insight into how review of portfolio-based information and performance improvement experiences can be coordinated with training programs to better enable program directors and licensing/certification bodies to encourage/monitor performance improvement efforts.

APPENDICES

Context

In October 2011 the Assessment Review Task Force (ARTF) of the Medical Council of Canada (MCC) submitted its report entitled “Recalibrating for the 21st Century” that outlined six issues facing the MCC and led to an equal number of recommendations. Chief among these for current purposes is that the ARTF expressed concern that the objectives emphasized in the blueprint of the MCC examinations needed “broadened and deepened” to better represent competencies beyond the medical expert role as articulated in the CanMEDS framework. The identification of this issue led to the recommendation that the MCC expand its assessment processes by (a) defining the knowledge and behaviours expected of physicians about to enter independent practice, (b) reviewing the adequacy of the blueprints for all MCC examinations, (c) revising said blueprints to improve fitness for purpose, and (d) exploring the development of new tools for any core competencies determined to not be assessable by current practices. That recommendation led to the launch of an extensive blueprinting project of which this report is intended to provide one component.

MCC Blueprint Project

The MCC Blueprint Project was approved and began in January 2012. At its root it is a multifaceted effort to perform a rigorous analysis of the knowledge, skills, and behaviours expected of all physicians licensed to practice medicine in Canada. Its aim is to guide content and strategy development to ensure that the program of assessment run by the MCC validly reflects its intended purpose. Pieces of the Blueprint Project include broad consultation with representatives of the medical schools and regulatory authorities in Canada, a large-scale national survey regarding the importance of including specific objectives on the MCC assessments, and consideration of empirical databases outlining the prevalence/severity of medical conditions impacting upon the health of Canadians. Consideration of the content included in any program of assessment, however, is insufficient as not all issues of relevance to the assessment of health professionals boil down to the simple question of what is assessed. In addition, it is important when deciding upon adoption of an assessment program to consider both the implicit messages sent to candidates and stakeholders and any unintended consequences of adopting a particular assessment strategy. This is Messick’s notion of consequential validity and requires a broader, scoping, review of not only current best practices in assessment, but also the alignment between current assessment systems and modern understanding of health professional practice. It is this aspect of the MCC Blueprint Project that the MCC’s Medical Education Assessment Advisory Committee was asked to address and which forms the basis of this report.

Medical Education Assessment Advisory Committee

The Medical Education Assessment Advisory Committee (MEAAC) was formed in late 2011 as a reinvention and repurposing of the group known as the Examination Development Advisory Committee (EDAC). It is an international panel of experts in the field of health professional education assessment who are external to the MCC. They are charged with advising the MCC directorate (Office of the CEO) and through them, the relevant committees of Council, on ongoing and future assessment processes that are necessary to enable MCC to meet its mission. MEAAC's specific roles and responsibilities in this regard are to: (a) provide MCC with an expert think tank to promote discussion, education and collaboration across committees, disciplines and organizations; (b) critically assess educational, psychological, and assessment literature in medicine, health care and other fields to enable the development or initiation of new approaches to medical assessment; (c) receive strategic plans and reports from internal and external working groups as determined by the MCC Research and Development office and the MCC directorate for the sake of providing critical reaction; and (d) prepare reports and make recommendations through the directorate on new approaches to assessment that will enable MCC to meet its mission.

Our task

The challenge put before MEAAC in Fall 2012 was a broad one: Consider the current state and future of medical education in Canada in the context of the current assessment practices of the MCC and of health professional education more broadly, and offer an opinion on what the MCC should be doing that it is not.

Our approach

We stimulated our conversations by reviewing three documents thought to contain important overviews of the current state of health professional assessment in Canada: (1) The Future of Medical Education in Canada (FMEC): A Collective Vision for MD Education, produced by the Association of Faculties of Medicine in Canada (AFMC); (2) Assessment in Postgraduate Medical Education: Trends and Issues in Assessment in the Workplace, produced as part of the environmental scan for the Future of Medical Education in Canada Postgraduate Project conducted by AFMC; and (3) Assessment Strategies within the Revised Maintenance of Certification Program, a white paper produced by the Royal College of Physicians and Surgeons of Canada. The group did not desire to simply re-produce these documents or the messages they contain under a new banner and, as such, while there will be some inevitable overlap in focus reflecting consistency of view, we merely recommend the Blueprint Project team to those additional supplementary perspectives.

From that foundation the members of MEAAC launched an iterative discussion about what they considered to be the key issues facing the MCC and related entities in the near future. This began with a face-to-face meeting in Ottawa in October of 2012 and continued over subsequent months via email and one-on-one telephone conversations. A draft version of this report was circulated at the end of February 2013, two weeks prior to the group having another face-to-face meeting in Ottawa. Revisions were incorporated in the month following that meeting and the final report was submitted on April 25th, 2013.

The three themes chosen for inclusion in this report should not be considered mutually exclusive as we believe drawing such boundaries would under-represent the complexity of medical education. Nor should they be considered comprehensive of all possible issues that could be addressed. The committee contemplated various issues that it did not have time to immediately explore in great depth and, as such, chose not to include them in this report apart from their being implicitly reflected in the three themes that are described. These include accountability (i.e., establishing an assessment system that is responsive to constantly evolving patient and community needs), convergence (or deliberate divergence) of assessment practices across discipline within profession and inter-professionally, the strengths/weaknesses inherent in trying to assess for particular approaches to practice (e.g., whether or not benefit is gained by using speeded examinations to assess automatic clinical reasoning processes as a supplement to assessing more deliberate consideration of patients cases), and the incorporation of patient-centred care into the assessment process as a critical outcome in some manner. These topics and others will be considered more extensively as the work of MEAAC continues.