

Medical Council of Canada Qualifying Examinations and performance in future practice

Les examens de certification du Conseil médical du Canada et le rendement futur des médecins

Elizabeth Wenghofer,¹ John Boulet²

¹School of Kinesiology and Health Sciences, Laurentian University; Division of Human Sciences, Northern Ontario School of Medicine, Ontario, Canada; ²National Board of Osteopathic Medical Examiners (NBOME); Uniformed Services University of the Health Sciences (USUHS), Bethesda, Maryland, USA

Correspondence to: Elizabeth Wenghofer, PhD, Full Professor, School of Kinesiology and Health Sciences, Laurentian University, Room SE205-G, 935 Ramsey Lake Road Sudbury, ON, P3E 2C6; phone: 705-675-1151 ext. 3925; email: ewenghofer@laurentian.ca

Published ahead of issue: June 7, 2022; published: Aug 26, 2022. CMEJ 2022, 13(4) Available at <https://doi.org/10.36834/cmej.73770>

© 2022 Wenghofer, Boulet; licensee Synergies Partners. This is an Open Journal Systems article distributed under the terms of the Creative Commons Attribution License. (<https://creativecommons.org/licenses/by-nc-nd/4.0>) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited.

Abstract

The purpose of medical licensing examinations is to protect the public from practitioners who do not have adequate knowledge, skills, and abilities to provide acceptable patient care, and therefore evaluating the validity of these examinations is a matter of accountability. Our objective was to discuss the Medical Council of Canada's Qualifying Examinations (MCCQEs) Part I (QE1) and Part II (QE2) in terms of how well they reflect future performance in practice.

We examined the supposition that satisfactory performance on the MCCQEs are important determinants of practice performance and, ultimately, patient outcomes. We examined the literature before the implementation of the QE2 (pre-1992), post QE2 but prior to the implementation of the new Blueprint (1992-2018), and post Blueprint (2018-present).

The literature suggests that MCCQE performance is predictive of future physician behaviours, that the relationship between examination performance and outcomes did not attenuate with practice experience, and that associations between examination performance and outcomes made sense clinically.

While the evidence suggests the MCC qualifying examinations measure the intended constructs and are predictive of future performance, the validity argument is never complete. As new competency requirements emerge, we will need to develop valid and reliable mechanisms for determining practice readiness in these areas.

Résumé

L'objectif des examens donnant lieu au titre de Licencié du Conseil médical du Canada est de protéger le public en garantissant que les praticiens possèdent les connaissances, les habiletés et les aptitudes nécessaires pour offrir des soins satisfaisants aux patients; par conséquent, l'évaluation de la validité de ces examens est une question de responsabilité. Notre objectif était de déterminer dans quelle mesure l'Examen d'aptitude du Conseil médical du Canada (EACMC), partie I, et l'EACMC, partie II reflètent le rendement futur des médecins dans leur pratique.

Nous avons examiné l'hypothèse selon laquelle des résultats satisfaisants aux EACMC sont des déterminants importants du rendement dans la pratique future et, ultimement, des résultats rapportés pour les patients. Nous avons examiné les écrits publiés avant l'introduction de l'EACMC, -partie II (avant 1992), post EACMC -partie II ci mais avant l'adoption du Plan directeur (1992-2018), ainsi que ceux publiés post adoption du Plan directeur (2018-présent).

La littérature suggère que la performance à l'EACMC permet de prédire les comportements futurs des médecins, que le rapport entre la performance à l'examen et les résultats dans la pratique perdure, et que les associations entre la performance à l'examen et les résultats sont liés sur le plan clinique.

Bien que les données probantes indiquent que les examens d'aptitude du CMC (EACMC) mesurent les concepts visés et permettent de prédire le rendement des médecins dans leur pratique future, la démarche de validité n'est pas complète. Au fur et à mesure que de nouvelles exigences en matière de compétences émergent, nous devons élaborer des mécanismes valides et fiables pour déterminer la capacité à exercer dans ces domaines

Introduction

In medicine and the other health professions, we assume that if we pick good learners and educate them well, this will lead to better patient care. We presuppose that licensure and certification processes identify those individuals who are ready for the next level of training or responsibility. Additionally, we need to be sure that such high stakes decisions are determined, from scores and pass/fail outcomes on reliable and valid assessments. With this in mind, evaluating our educational processes and associated assumptions is a requirement and thus is a matter of accountability to the public. We continually need to gauge how well our assumptions reflect reality.

Licensing examinations have been employed in many countries and in most, if not all, health professions.^{2,3} In combination with program accreditation mandates, they help ensure that individuals who provide care are adequately prepared to do so. In medicine, there has been considerable debate as to the value of licensure examinations.⁴⁻⁶ These examinations can be quite costly and force students to spend a great deal of time preparing, potentially contributing to excess stress and unwellness.⁷ Moreover, as assessments drive learning, licensure examination requirements can drive educational institutions to modify their curriculum, including the content and timing of coursework, to help increase their students' likelihood of passing. It has been argued that these curricular changes often impede student learning and can encourage educational institutions to deviate from their missions.⁸⁻¹⁰ Nevertheless, many countries and jurisdictions embrace licensure examinations, claiming that they help protect the public from practitioners who do not have adequate knowledge, skills, and abilities to provide acceptable patient care.^{11,12} Longitudinal approaches to assessment in medical education and certification show promise¹³ but a periodic and longitudinal approach to licensure examinations has yet to be explored.

In the United States and Canada, the successful completion of examinations, amongst other credentialing requirements, is needed to obtain an unrestricted license to practice medicine. As noted above, there are several arguments against the licensure process, or parts thereof, many directly related to the validity of the examinations. More specifically, there are claims that evidence linking examination performance to external criterion measures (e.g., practice as a physician) are lacking. Finally, studies of patient outcomes in countries with and without medical

licensure examinations have yielded few differences.¹⁴ Nevertheless, a host of investigations have linked licensure examination performance to later outcomes, including residency success, specialty board certification, patient outcomes, and physician disciplinary actions.¹⁵⁻²⁰ In spite of the fact that failing performance on a licensure examination restricts the individual from practicing medicine, it remains difficult to link individual practitioners to patient outcomes, and it is not possible or advisable to conduct randomized controlled trials (i.e., let those who fail a licensure examination obtain a licence so as to study their future practice outcomes). Nevertheless, the body of evidence supporting licensure and the validity of licensure examinations is reasonably strong, especially in Canada. In this position paper, we provide an overview of the evidence of the relationships found between examination performance and practice performance and patient outcomes.

This position paper was written to help the Assessment Innovation Task Force (AITF) develop recommendations for the future path of Medical Council of Canada (MCC) assessments. The purpose of this manuscript is to discuss the Medical Council of Canada's Qualifying Examinations (MCCQEs) Part I (QE1) and Part II (QE2) in terms of what they were designed/intended to measure and what they tell us about future practice. It is neither intended to be a systematic review of the literature nor a meta analysis of the existing evidence. Given that the QE1 and QE2 are part of Canada's medical licensure process, we provide a synthesis, based on the available literature, of how well these assessments identify individuals with the knowledge, skills, and abilities needed to provide quality patient care. This review of MCCQEs can both help inform potential future changes to the examination content and structure, and guide research efforts aimed at improving the medical licensure process.

Examination background

Until recently, the MCC administered two qualifying examinations: the MCC QE1 and QE2.²¹ From 1979 to 2018, the MCC Evaluating Examination (MCCQEE) was also offered to international medical graduates (IMGs). A passing score on this examination was required for IMGs to be eligible to attempt the QE1. As of 2019, all medical graduates, regardless of where they obtained their medical degree, were eligible to sit for the QE1.

The QE1 assesses the critical medical knowledge and clinical decision-making ability of a candidate at a level

expected of medical student who is completing their medical degree in Canada. It is currently a 1-day, computer-based, examination consisting of multiple-choice questions and a series of clinical decision making cases.²² The QE1 is a criterion-referenced examination. Those who meet or exceed the standard will pass the exam regardless of how well other candidates perform on it.

The purpose of the QE2 (no longer offered) was to assess the competence of candidates, specifically the knowledge, skills and attitudes essential for medical licensure in Canada, prior to entry into independent clinical practice. It was a 13-station (12 scored stations and 1 pilot) objective structured clinical examination (OSCE) that focused on the assessment of data acquisition skills, patient/physician interaction, problem solving and decision-making, and considerations for cultural communication. Qualified physician examiners in each station provided scores on the assessed dimensions using both checklists (i.e., physical examination) and rating scales (i.e., physician-patient interaction).²³ The MCC provides comprehensive examiner training comprised of both online education modules and compulsory day-of-exam orientation.

In 2018, the MCC launched a new comprehensive Blueprint that governed how content was distributed on the qualifying examinations.²⁴ Prior to this, content for QE1 was based on equal sections in Medicine, Surgery, Ob/Gyn, Pediatrics, Psychiatry and PHELO (Public Health, Ethics, Legal and Organizational aspects of medicine). This was true for both the multiple-choice question (MCQ) and clinical decision-making (CDM) parts of QE1. For QE2, prior to 2018, the content was specified by discipline (Medicine, Ob/Gyn, Pediatrics, Psychiatry, Surgery) and domain (Counseling/Education, History, Management, Physical Exam, Patient Interaction). With the implementation of the Blueprint, content is now focused on a combination of dimensions of care reflecting patient encounters (Health Promotion and Illness Prevention, Acute, Chronic, Psychosocial Aspects) and physician activities (Assessment/Diagnosis, Management, Communication and Professional Behaviours). The QE1 places emphasis on the physician activities of assessment/diagnosis and communication with the QE2 placing more emphasis on management followed by assessment/diagnosis.²⁴

Conceptual framework

Several validity frameworks can be used to categorize and synthesize evidence to support the psychometric adequacy of assessment scores or any decisions based on the

scores.²⁵⁻²⁹ All of these frameworks rely on evidence from numerous sources, including, amongst others, the specification of appropriate content domains(s), the accurate collection of examination scores, adherence to standardization protocols, adequate sampling of items (or cases) and, for high-stakes assessments, the use of defensible performance standards.³⁰ Of great importance, at least for medical licensure examinations, is the provision of evidence that performance on the examinations is related to performance in practice.

The MCC has been diligent in developing meaningful examination blueprints, allowing for the sampling of relevant content and providing some evidence to support content and construct validity. From a scoring perspective, regardless of the examination type, there are established quality assurance measures that ensure the accuracy of the scores. Finally, for both QE1 and QE2, defensible criterion-referenced standard-setting procedures are employed.^{22,23} Although the eligibility requirements, examination blueprints, and administration models (e.g., computer-based delivery of MCQs and CDMs) for the qualifying examinations have changed over time, the knowledge, skills, and abilities that are/were being measured (i.e., knowledge, application of knowledge, clinical reasoning, physical examination, history taking, communication) have remained fairly stable. As such, criterion-related evidence to support the validity of the qualifying examinations will not be confounded by small changes in the constructs being measured or exam administration modes.

While all of these strategies help ensure that the scores reflect the candidates' true abilities, they do not speak to the relationship between examination performance and performance as a practitioner. This relationship, a fundamental component of Kane's "extrapolation argument",³¹ is difficult to establish but is often considered the key validity criterion.

With the extrapolation argument in mind, we structure our discussion around the conceptual model presented in Tamblyn³² (derived from Kane²⁸), which describes the assumptions upon which qualifying examinations are based (Figure 1 - Original). We deviate slightly from her framework for ease of illustration. We suggest that prerequisites of competence (A) are those elements tested in the QE1 (i.e., knowledge as indicated through MCQs and clinical decision-making questions). Clinical competence (B) represents the endpoint of training programs or the result of the QE2.³² The results of both A and B are required to make a decision regarding licensure and practice

readiness. We examine the assumption that satisfactory performance at A and/or B are important determinants of practice performance (C) and ultimately patient outcomes (D).³² For the purposes of our discussion, practice performance (C) reflects activities related to physician practice behaviours associated with processes of care where the unit of analysis is the physician (e.g., communication skills, professionalism); patient outcomes (D) refer to clinical measures at the patient level (Figure 2 - Modified). Additionally, we overlay elements of the MCC Blueprint²⁴ to determine if the examinations designed to determine competence in specified physician activities and patient encounters are reflected in practice performance and patient outcomes (Figure 2 – Modified).

In building an argument to support the validity of the MCCQE exam scores (and decisions based on the exam scores), it should be noted that the qualifying examinations are/were intended to identify a minimum standard for safe

practice, not necessarily to differentiate levels of good performance. Moreover, individuals who do not receive the Licentiate of the Medical Council of Canada (LMCC) are not eligible to practice medicine in Canada. It can also be difficult to attribute outcomes to specific physicians because they often work in teams. Finally, as Tamblyn et al. 2011 noted,³³ associations between examination scores and patient outcomes often exist even though the reliability of the exam scores varies and propose that the magnitudes of these associations would increase if the reliability of exam scores improved. With these issues in mind, establishing relationships between examination scores (or decisions) and future performance as a physician, which can be attenuated because lower ability candidates may never receive licenses to practice, is difficult. However, when such relationships are found, they provide strong evidence that the examinations measure the knowledge, skills, and abilities needed to provide quality patient care.

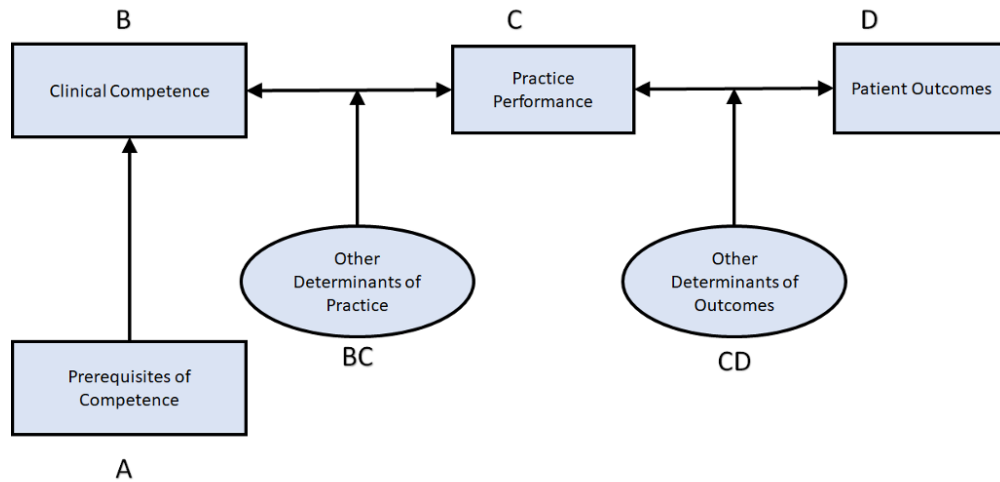


Figure 1. Interpretation of Licensing Examination scores: the assumptions (as found in Tamblyn^{32,p. 203})

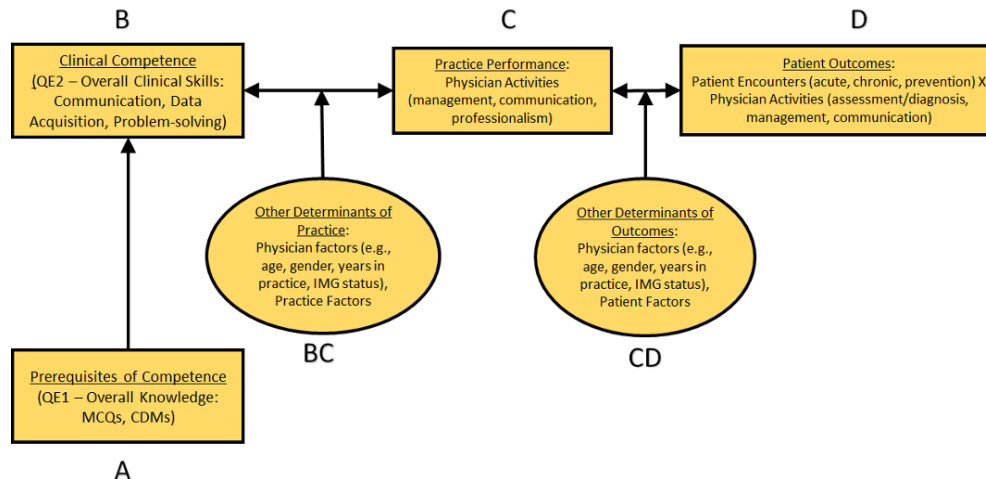


Figure 2. Conceptual Framework for Examining the Relationship between MCCQE1 and MMCQE2 Scores and Practice Performance and Patient Outcomes with Overlay of the MCC Blueprint (Modified from Tamblyn^{32,p. 203})

Results

We limit our interpretation of the literature to studies specifically focusing on MCC examinations. A broader overview of evidence to support the validity of medical licensing examinations was provided in the introduction. We divide the literature concerning the MCC QEs into three periods in time (i) before the implementation of the QE2 (pre-1992), (ii) post QE2 but prior to the implementation of the new Blueprint (1992-2018) and (iii) post Blueprint (2018-present). The majority of the QE1 and QE2 outcome studies were completed through a large-scale research program by Dr. Robyn Tamblyn and colleagues in period two and focused on the results of the first three years of examinations post QE2 implementation (1993-1996) with a follow-up period to 2007.³⁴⁻³⁹ Preliminary findings from this research program were summarized and presented with recommendations to the MCC.³³ Several other studies round out the evidence available for this period.^{40,41} Very few published studies focused on data prior to 1992^{42,43} or after the implementation of the new Blueprint.⁴⁴

The majority of studies quantified the relationship between examination scores and patient outcomes (Figure 2 - D), with fewer studies focusing on practice performance (Figure 2 - C). Patient outcome (Figure 2 -D) indicators focused on mammography screening rates,^{39,42} continuity of care indices,⁴² management of hypertension³⁶ and asthma,³⁷ and appropriate prescribing (i.e., appropriate prescribing of antibiotics,^{38,43} opioids,⁴⁴ disease-specific vs symptom relief prescribing⁴² and contraindicated prescribing rates).⁴² Outcome studies focused on primary care physicians as well as patient outcomes of physicians across specialty areas. Practice performance indicators (Figure 2 - C) focused on public complaints to regulatory authorities^{34,44} and performance on in-practice physician assessments.³⁵ Two studies focused on the relationship with qualifying examinations before entry to independent practice evaluating QE predictive association between success on certification examinations in IMGs⁴¹ and diagnostic accuracy and reaction time on case-based exercise.⁴⁰ Practice performance indicator studies focused on physicians across all speciality areas.

The evidence across indicators points in the same direction; those examinees who perform better on qualifying examinations are more likely to perform better based on practice performance and patient outcome indicators. No studies indicated a negative relationship. Only one study found no significant relationship between

examination scores (QE1) and patient outcomes (antibiotic prescribing for viral respiratory infections).⁴³ Regardless of whether examination performance was evaluated by the number of pass attempts,⁴⁴ increases in score per standard deviation^{34,42} or quartile range,^{34,35} better examination performance was associated with improved outcomes. All studies attempted to correct for other determinants of practice performance (Figure 2 – BC) and patient outcomes (Figure 2 – CD). In some studies,⁴⁴ these covariates were more strongly associated with the indicator of interest than were examination outcomes. Nonetheless, both QE1 and QE2 were consistently independently predictive of practice performance and patient outcomes.

Although there are positive predictive relationships between exam performance and relevant criterion measures, they do not tell us if individuals who pass the qualifying examinations are 'safe to practice.' There are two reasons for this: first, individuals who fail the examinations do not obtain licensure and thus have no practice or patient outcomes; second, none of the practice performance or patient outcomes evaluated to date reflect 'safe' vs. 'not safe.' Thus, we can only look at gradations in scores or examine those who take more than one attempt to pass and see how these measures relate to various outcomes. There is some evidence that practice and patient outcomes of individuals who required multiple attempts to pass are poorer than those who pass on their first attempt.^{41,44}

Although associations were found between outcome indicators and examination scores in every study save one, the associations were not necessarily found with every examination component evaluated. Nonetheless, both significantly positive and statistically insignificant associations were “sensible” based on the clinical competencies needed for good practice per specific indicator. By sensible we mean that the competencies measured by the examination were congruent with those required for a good patient outcome or practice performance. For example, the QE2 communication subscore was associated with indicators that required or were related to communication skills. Tamblyn et al. 2007 found that examinees who scored in the lowest quartile on QE1 and QE2 overall were more likely to have higher rates of complaints made to regulatory authorities that required investigations.³⁴ The strongest relationship existed between the performance on the communications portion of the QE2 and rates of communication-based complaints, whereas no significant relationship existed between

communication complaint rates and score on the data acquisition or problem-solving score on the QE2. Given that communication issues are often at the core of complaints, including those associated with clinical quality, this is a logical and expected association. De Champlain et al. 2020 found a similar association between examination scores and complaints but only with QE1 and not QE2.⁴⁴ However, their focus was on all complaints regardless of their nature or seriousness or those requiring further investigation. Another example of this logical association is found in Meguerditchian et al. 2012, where physicians whose patients had better mammography screening rates performed better on the communication aspect of the QE2.³⁹ However, no association was found for this indicator with QE1 score overall, QE2 score overall, or QE2 data acquisition score. It makes sense that physicians with better communication skills can engage their patients in preventive screening more effectively than those who cannot. Sherbino et al. 2012 evaluated examinees immediately after completing the QE2 for diagnostic accuracy during a computer-based case review.⁴⁰ They found that accuracy was higher for those with higher overall scores on the problem-solving portion of the QE1, but there was no association with the communication portion, data acquisition sub-score, or QE2 score overall.

Clinically sensible relationships crosscut the study evidence. Practice or outcome indicators that require strong communication skills are associated with communication sub-scores on the QE2 but associated with the QE1 or other aspects of the QE2 to lesser degrees. Those indicators requiring a solid knowledge base and decision-making skills are more strongly associated with the QE1 but less so with the communication portion of the QE2, and so forth, suggesting that the different components of these exams are evaluating different competencies as designed. Further supporting this conclusion, the correlations between QE2 sub-scores and the correlations between QE1 and QE2 results were not strong.^{34,36} Likewise, there was no relationship found between the QE1 examination and the Quebec family medicine certification examination (QLEX)⁴² and the performance component on the CFPC certification examination completed,^{2,41,45} which is as expected given that the QE1 evaluates cognitive competencies (i.e., knowledge) rather than communication or clinical skills. Lastly, an important finding that crosscuts studies that evaluated interactions between time in-practice and examination scores indicated that the relationship

between examination performance and outcomes did not attenuate with practice experience.^{36,38,39,42}

Discussion

The validity and utility of health professions licensing examination scores have been studied by many organizations, sometimes with mixed results.^{2,4,14,46} Ideally, the associated examinations would screen out those candidates with inadequate knowledge, skills, and abilities. The accuracy of these decisions, which depends on the scores' reliability and validity, is difficult to establish. However, for those making it through the licensure examination process, one would expect that the examination scores, if measuring the appropriate constructs, would be related to future practice metrics, including patient outcomes and disciplinary actions (or complaints). To the extent that these associations exist and care has been taken to set criterion-referenced standards, the validity of the examination(s) is supported.

Returning to the conceptual framework, our review of the literature supports that the QE1, representing a measure of prerequisites of competence (Figure 2 - A) and the QE2, representing a measure of clinical competence (Figure 2 - B), are positively associated with practice performance (Figure 2 - C) and patient outcomes (Figure 2 - D). Given that the purpose of qualifying examinations is to establish the minimum standard for a physician entering independent practice rather than predicting future practice,⁴⁷ the finding that examination scores are still predictive of outcomes seven to ten years later^{36,38,39,42} provides further support for the validity of the examinations.

It is worth noting that much of the literature did not look at the various component sub-component scores of the QE1 or QE2, but rather focused on overall examination performance. In particular, the QE2 data acquisition sub-scores and problem-solving sub-scores were often not evaluated, and were generally less predictive than the overall QE2 scores or QE2 communication sub-scores. However, the associations found (and those not found) made clinical sense given the indicators. So, although the current evidence makes it difficult to say what the data acquisition sub-scores and problem-solving sub-scores on the QE2 indicate for practice, we can say that they contribute to the overall picture, and in a clinically logical way. While more work needs to be done looking at these sub-scores, the QE2 is no longer administered. As such, and without a replacement of this examination, it would be

informative to contrast future outcomes (e.g., residency performance, complaints) for those who did and did not take the QE2. With the QE1, many studies focused on overall scores did not specifically look at MCQ or clinical decision-making sub-scores; future work in this area, at least from a validity standpoint, would be informative.

From the current evidence, we cannot comment on whether examination performance in the areas reflecting the subtle aspects of the physician activities outlined in the MCC Blueprint²⁴ is related to practicing physician performance in these areas. In particular, we refer to performance on intrinsic competencies associated with professional behaviour, including ethics, empathy, self-awareness, leadership, etc. With the exception of regulatory complaints,^{34,44} which could be considered an indicator of professionalism, at least to some degree, studies rarely concentrate on non-clinical competencies that are deemed essential for entry to practice. Interestingly, the only study which specified professionalism as an outcome was Tamblyn et al. 2007,³⁴ who did not find an association between professionalism-based complaints and QE1 or QE2 scores. This is an important area for future study as practicing physicians often identify practice challenges associated with these intrinsic competencies, yet very few recognize them as learning needs.⁴⁸ Evaluating how to most effectively determine practice readiness in these essential intrinsic competencies is imperative for safe, respectful patient-centred care. Lastly, with the cancellation of the QE2 with no current replacement, it is possible that physicians entering practice may have diminished communication skills which could result in increased complaints to regulatory authorities. It will be important to monitor this the future until a replacement for the QE2 is found. Alternatively, residency programs may need to increase emphasis on communication competency.

Summary

While primarily focused on the extrapolation component of Kane's validity argument, the review of the available literature suggests that performance on the MCCQEs is predictive of future physician behaviours. While evidence suggests that MCCQEs are measuring the intended constructs and are predictive of future performance, the validity argument is never complete. The predictive relationships that have been established to date may become irrelevant as medicine evolves. Although there is evidence to suggest that the association between QE

performance and practice has existed over an extended timeframe, which is an important aspect of validity, this cannot tell us if we are focusing on the right aspects of competence and performance to determine practice readiness in the twenty-first century. For example, the recent pandemic has highlighted the importance of performance in virtual care as an essential competency; thus, we must now consider how to determine practice readiness with this and other emerging competency areas (e.g., use point of care ultrasound). As the practice of medicine changes, forcing changes to the knowledge, skills, and abilities measured as part of the licensure process, additional validation work is necessary.

Conflicts of Interest: Dr. Wenghofer has no conflicts of interest to declare, financial or otherwise. Dr. Boulet was the interim Director of Assessment at the Medical Council of Canada, directly overseeing the exam delivery and operations related to the MCCQE Part I and NAC examinations.

References

1. Wenghofer EF. Research in medical regulation: an active demonstration of accountability. *J Med Reg.* 2015;101(3):13-7. <https://doi.org/10.30770/2572-1852-101.3.13>
2. Swanson DB, Roberts TE. Trends in national licensing examinations in medicine. *Med ed.* 2016;50(1):101-14. <https://doi.org/10.1111/medu.12810>
3. Bobos P, Pouliopoulou DV, Harriss A, Sadi J, Rushton A, MacDermid JC. A systematic review and meta-analysis of measurement properties of objective structured clinical examinations used in physical therapy licensure and a structured review of licensure practices in countries with well-developed regulation systems. *PLoS one.* 2021;16(8):e0255696. <https://doi.org/10.1371/journal.pone.0255696>
4. Archer J, Lynn N, Coombes L, Roberts M, Gale T, Regan de Bere S. The medical licensing examination debate. *Regul Gov.* 2017;11(3):315-22. <https://doi.org/10.1111/rego.12118>
5. Archer J, Lynn N, Roberts M, Coombes L, Gale T, de Regand Bere S. *A systematic review on the impact of licensing examinations for doctors in countries comparable to the UK.* Plymouth: Plymouth University. 2015.
6. Schuwirth L. National licensing examinations, not without dilemmas. *Med ed.* 2016;50(1):15-7. <https://doi.org/10.1111/medu.12891>
7. Tschlis JT, Del Re AM, Carmody JB. The past, present, and future of the united states medical licensing examination step 2 clinical skills examination. *Cureus.* 2021;13(8). <https://doi.org/10.7759/cureus.17157>
8. Wormald BW, Schoeman S, Somasunderam A, Penn M. Assessment drives learning: an unavoidable truth? *Anat sci educ.* 2009;2(5):199-204. <https://doi.org/10.1002/ase.102>
9. Murray DJ, Boulet JR. Anesthesiology Board Certification Changes: A Real-time Example of "Assessment Drives Learning." *Anesthesiology.* 2018;128(4):704-6. <https://doi.org/10.1097/ALN.0000000000002086>
10. Norman G, Neville A, Blake JM, Mueller B. Assessment steers learning down the right road: impact of progress testing on licensing examination performance. *Med teach.*

- 2010;32(6):496-9.
<https://doi.org/10.3109/0142159X.2010.486063>
11. Boulet J, van Zanten M. Ensuring high-quality patient care: the role of accreditation, licensure, specialty certification and revalidation in medicine. *Med ed* 2014; 48: 75-86
<https://doi.org/10.1111/medu.12286>
 12. Schuwirth L. The need for national licensing examinations. *Med ed*. 2007;41(11):1022-3. <https://doi.org/10.1111/j.1365-2923.2007.02856.x>
 13. Boulet JR, Durning SJ. What we measure... and what we should measure in *Med ed*. 2019;53(1):86-94.
<https://doi.org/10.1111/medu.13652>
 14. Archer J, Lynn N, Coombes L, et al. The impact of large scale licensing examinations in highly developed countries: a systematic review. *BMC med ed*. 2016;16(1):212-212.
<https://doi.org/10.1186/s12909-016-0729-7>
 15. Roberts WL, Gross GA, Gimpel JR, et al. An investigation of the relationship between COMLEX-USA licensure examination performance and state licensing board disciplinary actions. *Acad Med*. 2020;95(6):925-30.
<https://doi.org/10.1097/ACM.0000000000003046>
 16. Cuddy MM, Young A, Gelman A, et al. Exploring the relationships between USMLE performance and disciplinary action in practice: a validity study of score inferences from a licensure examination. *Acad Med*. 2017;92(12):1780-5.
<https://doi.org/10.1097/ACM.0000000000001747>
 17. Norcini JJ, Boulet JR, Opalek A, Dauphinee WD. The relationship between licensing examination performance and the outcomes of care by international medical school graduates. *Acad Med*. 2014;89(8):1157-62.
<https://doi.org/10.1097/ACM.0000000000000310>
 18. Hudson KM, Feinberg G, Hempstead L, Zipp C, Gimpel JR, Wang Y. Association between performance on COMLEX-USA and the American College of Osteopathic Family Physicians in-service examination. *J Grad Med Ed*. 2018;10(5):543-7.
<https://doi.org/10.4300/JGME-D-17-00997.1>
 19. Arnhart KL, Cuddy MM, Johnson D, Barone MA, Young A. Multiple United States medical licensing examination attempts and the estimated risk of disciplinary actions among graduates of US and Canadian medical schools. *Acad Med*. 2021;96(9):1319-23.
<https://doi.org/10.1097/ACM.0000000000004210>
 20. Patel MD, Heitkamp DE, Jordan SG. Re: The relationship between US medical licensing examination step scores and ABR core examination outcome and performance: a multi-institutional study. *J Am Coll Radiol*. 2021;18(1):8-9.
<https://doi.org/10.1016/j.jacr.2020.10.018>
 21. Medical Council of Canada. *The MCC suspends the delivery of the MCCQE Part II* [Internet]. Available from: <https://mcc.ca/news/mcc-suspends-delivery-of-mccqe-part-ii/> [Accessed on Nov 7, 2021].
 22. Medical Council of Canada. *Medical Council of Canada Qualifying Examination (MCCQE) Part I Annual Technical Report 2019* [Internet]. 2019 Apr. Available from: <https://mcc.ca/media/MCCQE-Part-I-Annual-Technical-Report-2019.pdf> [Accessed on Apr 1, 2022].
 23. Medical Council of Canada. *Medical Council of Canada Qualifying Examination (MCCQE) Part II Annual Technical Report 2018* [Internet]. 2018 Oct. Available from: <https://www.mcc.ca/media/MCCQE-Part-II-Technical-Report-Oct-2018.pdf> [Accessed on Apr 1, 2022].
 24. Medical Council of Canada. *Blueprint Project: Examinations Blueprint and Content Specifications - Final Version* [Internet]. 2014 Sep. Available from: <https://mcc.ca/media/Blueprint-Report.pdf> [Accessed on Nov 7, 2021].
 25. Pugh DM, Wood TJ, Boulet JR. Assessing procedural competence: validity considerations. *Simul Healthc* 2015;10(5):288-94.
<https://doi.org/10.1097/SIH.000000000000101>
 26. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med ed* 2015; 49: 560-575
<https://doi.org/10.1111/medu.12678>
 27. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 2014.
 28. Kane MT. Validating interpretive arguments for licensure and certification examinations. *Eval Health Prof*. 1994;17(2):133-59. <https://doi.org/10.1177/016327879401700202>
 29. Kane M. Validating high-stakes testing programs. *Educ Meas: Issues Pract*. 2002;21(1):31-41.
<https://doi.org/10.1111/j.1745-3992.2002.tb00083.x>
 30. Clauser B, Margolis M, Swanson D. Issues of validity and reliability for assessments in medical education. In: Holmboe ES, Durning SJ, Hawkins RE, editors. *Practical guide to the evaluation of clinical competence*. Second edition. Philadelphia, PA: Elsevier; 2018. p. 22-36.
 31. Kane M. The argument-based approach to validation. *School Psych Rev*. 2013;42(4):448-57.
<https://doi.org/10.1080/02796015.2013.12087465>
 32. Tamblyn R. Is the public being protected? prevention of suboptimal medical practice through training programs and credentialing examinations. *Eval Health Prof*. 1994;17(2):198-221. <https://doi.org/10.1177/016327879401700205>
 33. Tamblyn R, Abrahamowicz M, Bartlett G, et al. *Executive summary. Medical Council of Canada: The Quebec-Ontario Follow-up Study of the Association between Scores Achieved on the MCCQE Part II Examination and Performance in Clinical Practice* [Internet]. 2011. Available from: https://mcc.ca/media/Tamblyn_Score-Association_MCCQE-Part-II_Clinical-Practice-Performance_2009.pdf [Accessed on Jun 5, 2021].
 34. Tamblyn R, Abrahamowicz M, Dauphinee D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *JAMA*. 2007 Sep 5;298(9):993-1001. <https://doi.org/10.1001/JAMA.298.9.993>
 35. Wenghofer E, Klass D, Abrahamowicz M, et al. Doctor scores on national qualifying examinations predict quality of care in future practice. *Med ed*. 2009;43(12):1166-73.
<https://doi.org/10.1111/j.1365-2923.2009.03534.x>
 36. Tamblyn R, Abrahamowicz M, Dauphinee D, et al. Influence of physicians' management and communication ability on patients' persistence with antihypertensive medication. *Arch Intern Med*. 2010;170(12):1064-72.
<https://doi.org/10.1001/archinternmed.2010.167>
 37. Kawasumi Y, Ernst P, Abrahamowicz M, Tamblyn R. Association between physician competence at licensure and the quality of asthma management among patients with out-of-control asthma. *Arch Intern Med*. 2011;171(14):1292-4.
<https://doi.org/10.1001/archinternmed.2011.319>
 38. Cadieux G, Abrahamowicz M, Dauphinee D, Tamblyn R. Are physicians with better clinical skills on licensing examinations less likely to prescribe antibiotics for viral respiratory infections in ambulatory care settings? *Med care*. 2011;156-65.
<https://doi.org/10.1097/MLR.0b013e3182028c1a>

39. Meguerditchian A-N, Dauphinee D, Girard N, et al. Do physician communication skills influence screening mammography utilization? *BMC*. 2012;12(1):1-8. <https://doi.org/10.1186/1472-6963-12-219>
40. Sherbino J, Dore KL, Wood TJ, et al. The relationship between response time and diagnostic accuracy. *Acad Med*. 2012;87(6):785-91. <https://doi.org/10.1097/ACM.0b013e318253acbd>
41. De Champlain AF, Streefkerk C, Roy M, Tian F, Qin S, Brailovsky C. Predicting family medicine specialty certification status using standardized measures: for a sample of international medical graduates engaged in a practice-ready assessment pathway to provisional licensure. *J Med Reg*. 2014;100(4):8-16. <https://doi.org/10.30770/2572-1852-100.4.8>
42. Tamblyn R, Abrahamowicz M, Dauphinee WD, et al. Association between licensure examination scores and practice in primary care. *JAMA*. 2002;288(23):3019-26. <https://doi.org/10.1001/JAMA.288.23.3019>
43. Cadieux G, Tamblyn R, Dauphinee D, Libman M. Predictors of inappropriate antibiotic prescribing among primary care physicians. *CMAJ*. 2007;177(8):877-83. <https://doi.org/10.1503/CMAJ.070151>
44. De Champlain AF, Ashworth N, Kain N, Qin S, Wiebe D, Tian F. Does pass/fail on medical licensing exams predict future physician performance in practice? a longitudinal cohort study of Alberta physicians *J Med Reg*. 2020;106(4):17-26. <https://doi.org/10.30770/2572-1852-106.4.17>
45. Kane MT, Crooks TJ, Cohen AS. Designing and evaluating standard-setting procedures for licensure and certification tests. *Adv Health Sci Educ*. 1999;4(3):195-207. <https://doi.org/10.1023/A:1009849528247>
46. Price T, Lynn N, Coombes L, Roberts M, Gale T, de Bere SR, et al. The international landscape of medical licensing examinations: a typology derived from a systematic review. *International J health policy and management*. 2018;7(9):782. <https://doi.org/10.15171/ijhpm.2018.32>
47. Breithaupt K. *Medical Licensure Testing White Paper for the Assessment Review Task Force of the Medical Council of Canada*. 2011.
48. McConnell M, Gu A, Arshad A, Mokhtari A, Azzam K. An innovative approach to identifying learning needs for intrinsic CanMEDS roles in continuing professional development. *Med ed online*. 2018;23(1):1497374. <https://doi.org/10.1080/10872981.2018.1497374>