

Medical Council
of Canada
Qualifying
Examination
(MCCQE) Part II

SPRING

Spring 2018
MCCQE Part II
Annual
Technical
Report



MEDICAL COUNCIL
OF CANADA

LE CONSEIL MÉDICAL
DU CANADA

Candidate

Admin: 1
Track: C
Start: 06

TABLE OF CONTENTS

Preface	4
1. Overview of the MCCQE Part II	4
2. Exam development	5
2.1 Exam specifications.....	5
2.2 Case writing.....	6
2.3 Case review and piloting	7
2.4 Exam format	7
2.5 Content validity.....	8
2.6 Exam scoring.....	9
3. Candidate orientation, exam administration and training	10
3.1 Candidate registration and eligibility	10
3.2 Candidate orientation	10
3.3 Exam administration.....	11
3.4 Exam administrative staff	11
3.5 Standardized Patient training	12
3.6 Deputy Registrars.....	12
3.7 Chief Examiners	13
3.8 Physician Examiner recruitment and training.....	13
4. MCCQE Part II scoring	15
4.1 Standard quality assurance and quality control procedures	15
4.2 Exam result approval.....	16
4.3 Exam result reporting	16
4.3.1 Scale score.....	17
4.3.2 Pass/fail status	19
4.3.3 Domain subscores.....	20
5. Psychometric results.....	20
5.1 Scale scores	22
5.2 Estimates of reliability and classification decisions	24
5.2.1 Cronbach's Alpha	24
5.2.2 Standard error of measurement	24
5.2.3 Decision consistency and decision accuracy	24
5.3 OSCE station statistics.....	25
5.4 Examiner analyses	27
5.5 Domain subscore profiles.....	28
5.6 Historical pass rates	29
6. References.....	33
Appendix A: Quality control – MCCQE Part II results.....	34
Appendix B: Patient interaction rating scale items	39
Appendix C: Example SOR and SFR	43

LIST OF TABLES AND FIGURES

Table 1: Exam specifications for the MCCQE Part II	5
Table 2: Domain definitions for the MCCQE Part II (domain definitions)	6
Table 3: Sampling of OSCE exam specifications for each test form.....	8
Table 4: Sampling of OSCE subscores for each MCCQE Part II test form.....	9
Table 5: MCCQE Part II reporting domains.....	9
Table 6: PE recruitment requirements.....	13
Table 7: Number and percentage of candidates for the MCCQE Part II by PGT group	21
Table 8: Summary statistics of scale scores by PGT group.....	22
Figure 1: Scale score distribution for May 2018 for total candidates and by PGT group.....	23
Figure 2: Scale score distribution for May 2018 for CMG-CPG first-time test takers, by PGT group	23
Table 9: Reliability estimates, SEM, decision consistency and accuracy for May 2018	25
Table 10: Summary statistics for OSCE stations for May 2018	26
Figure 3: Domain subscore for May2018 MCCQE Part II candidates	29
Table 11: 2013 – May 2018 pass rates by total reference group.....	30
Table 12: Spring 2013 – May 2018 pass rates by total reference group	31
Table 13: Fall 2013 – 2017 pass rates by total reference group	32

PREFACE

This report summarizes the exam development, exam administration, scoring and psychometric activities of the Medical Council of Canada Qualifying Examination (MCCQE) Part II and candidate performance on the exam in May 2018. Sections 1 to 4 describe the exam's purpose, format, content development, administration, scoring and score reporting. These sections also provide validity evidence in support of score interpretation, reliability and errors of measurement, and other psychometric characteristics. Section 5 summarizes candidate performances for the May 2018 administration and includes historical data for reference purposes. The report is intended to serve as technical documentation and reference material for members of the Central Examination Committee (CEC), test committee members and Medical Council of Canada (MCC) staff.

1. OVERVIEW OF THE MCCQE PART II

The purpose of the Medical Council of Canada Qualifying Examination (MCCQE) Part II is 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.

The exam is a 13-station Objective Structured Clinical Examination (OSCE) that focuses on the assessment of data gathering, physical exam skills, communication skills, and considerations of cultural communication, legal, ethical and organizational (C²LEO) aspects of the practice of medicine. The exam consists of a series of 12 stations that count towards the candidate's total score and one pilot station that does not count. At each station, a brief, written statement introduces a clinical problem and directs the candidate to appropriately examine a Standardized (simulated) Patient (SP) and to perform activities such as obtaining a focused history, conducting a focused physical exam or assessing and addressing the patient's issues. Candidates may be asked to answer specific questions related to the patient, interpret X-rays or the results of other investigations, make a diagnosis and/or write admission orders. The MCCQE Part II includes problems in medicine, pediatrics, obstetrics and gynecology, preventive medicine and community health, psychiatry, surgery and similar disciplines considered essential for competence in general medicine and health care.

Candidates are eligible to challenge the MCCQE Part II after their medical degree has been successfully source verified, they have passed the MCCQE Part I and successfully completed a minimum of 12 months of Postgraduate clinical medical Training (PGT) or osteopathic

postgraduate training on or before the deadline of June 30 for the May exam of the same calendar year or December 31 for the October exam of the same calendar year. The exam is offered twice per year, once in May and again in October. It is scored by Physician Examiners (PEs) and the performance standard, also known as a pass score, reflects a candidate who is minimally competent to enter independent practice.

The Centralized Examination Committee (CEC) is responsible for the overall content and quality of the examinations and the approval of results.

2. EXAM DEVELOPMENT

2.1 EXAM SPECIFICATIONS

Exam specifications were developed for the MCCQE Part II and approved by the OSCE Test Committee in 2004. The specifications seen below outline the domains and disciplines of the OSCE stations to ensure that similar content is measured on each of the exam forms. Creating an exam form that meets exam specifications ensures that candidates are measured on similar content and can be compared fairly from one exam to the next.

Table 1 provides the exam specifications and constraints including domain and discipline for the MCCQE Part II.

Table 1: Exam specifications for the MCCQE Part II

		Stations
<i>Domain</i>	Counselling/education	2
	History	4 or 5
	Management	2
	Physical exam	3 or 4
	History + PE when possible	(2)*
<i>Discipline</i>	Medicine	3
	OBGYN	2
	Pediatrics	2
	Psychiatry	2
	Surgery	3

* When possible, two combined history-taking and physical exam stations

Patient interaction is not a separate domain but is integrated within most stations, except for some acute care stations.

Table 2 provides the definition for each domain outlined in Table 1 along with a description of what candidates are given credit for in each domain.

Table 2: Domain definitions for the MCCQE Part II (domain definitions)

Counselling/ Education	History	Management (ER)	Physical exam	Patient interaction
Responding to a patient problem that includes a counselling challenge	Taking an organized and focused history	Managing an acute patient problem or trauma	Conducting a focused physical exam	
Credit is given for: <ul style="list-style-type: none"> Eliciting key information about the patient's problem Providing information Understanding how the patient perceives the problem Advising the patient Recommending follow-up 	Credit is given for: <ul style="list-style-type: none"> Demonstrating an understanding of the most likely differential diagnosis and of the urgency of the problem 	Credit is given for: <ul style="list-style-type: none"> Setting clinical priorities Eliciting critical information Conducting physical diagnostic maneuvers Ordering investigations Starting initial treatment Suggesting follow-up Nurses are sometimes available to order tests or perform procedures requested by candidates 	Credit is given for: <ul style="list-style-type: none"> Conducting relevant physical diagnostic maneuvers Recognizing the level of urgency of the problem Performing maneuvers satisfactorily Reporting negative or positive findings Standardized patients are used, not models 	<ul style="list-style-type: none"> Patient interaction is not treated as a separate domain, but is integrated within most stations Some acute care stations are the exception

2.2 CASE WRITING

Content for the MCCQE Part II is developed by panels of physicians along with experts in medical education and assessment from across Canada. A thorough process is followed to ensure that issues related to content, feasibility, authenticity and reproducibility are addressed early in case development. MCCQE Part II case authors develop content that reflects the MCC's expectations of competent physicians in independent practice in Canada.

Case authors create stations that deal with data gathering, diagnostic, clinical problem solving and the principles of management that are applicable, in whole or in part, to the range of clinical situations commonly encountered by physicians.

Case authors first develop the information given to candidates prior to entering an OSCE station. This includes the required clinical task and the time allotted to candidates to perform this task.

The authors then develop instruments to score the candidate's performance. Scoring instruments may consist of a checklist, rating scales, and oral or written questions. Authors also create any supporting materials, such as an X-ray or a patient chart, relating to tasks that candidates might have to complete prior to seeing the patient and/or during the patient encounter.

In addition, case authors compile information for SPs and SP trainers. To ensure that SPs and SP trainers understand the problem from the patient's perspective, case authors provide extensive information on the presenting problem, the patient's behaviour, appearance and affect, the physical findings to be portrayed as well as their appropriate medical and social history. Specific instructions about the room set-up, props, and required equipment are also provided by the authors.

2.3 CASE REVIEW AND PILOTING

Case development is an iterative process that requires careful thought, review, and revision. Once a case is created, its content is reviewed by clinical colleagues, test committee members, and MCC staff to improve the quality of the case and identify potential problems early in the process. For instance, role-playing the case allows authors to identify necessary information that may be missing from the SP script. Following an extensive review process, a case is piloted during a live exam to determine how well it functions both logistically and psychometrically.

The MCC is currently developing stations required to meet the domains outlined in the new Blueprint. In May 2018, 14 new cases were piloted. More information about MCC's new Blueprint can be found in a report called *Blueprint Project: Qualifying Examinations Blueprint and Content Specifications* (Medical Council of Canada, 2014).

2.4 EXAM FORMAT

The MCCQE Part II is a two-day examination that consists of a series of clinical stations, including pilot stations that do not count towards a candidate's total score. In May 2018, candidates attempting the MCCQE Part II completed an exam form composed of eight scored, 10-minute encounter stations on Day 1 (Saturday), and four scored couplet stations and one 14-minute pilot station on Day 2 (Sunday). Couplet stations consist of an encounter with a patient, either preceded by a task such as reading a chart or followed by a written assessment relating to the patient encounter. In May 2018, the couplet stations consisted of six-minute encounters and six-minute tasks.

At each station, a brief written statement introduces a clinical problem and directs the candidate to appropriately examine an SP as well as perform activities such as obtaining a focused history, conducting a focused physical exam or assessing and addressing the patient's issues.

Standardized administration, PE/SP training, and analytic procedures are followed to ensure that the candidate’s scores are comparable over time. Detailed steps of the analytic procedures are described in Appendix A: Quality Control – MCCQE Part II Results (MCC, 2017).

2.5 CONTENT VALIDITY

Measuring how well an exam form matches the exam specifications is one piece of validity evidence supporting valid score interpretations and arguments for the intended purpose of the exam (Kane, 2006, 2013). This section highlights the exam specifications and how well each exam form measures the exam specifications.

The MCCQE Part II Test Committee works with MCC staff to select and approve the OSCE stations for a given exam form. The exam forms are drafted by the MCCQE Part II Test Development Officer in accordance with exam specification criteria. The MCCQE Part II Test Committee then reviews the exam forms, including individual OSCE cases, to ensure that exam specification criteria have been met and that content is at the appropriate assessment level. This specifically refers to the knowledge, skills, and attitudes essential for medical licensure in Canada prior to entry into independent clinical practice. The MCCQE Part II Test Committee approves the final version of the content for each exam form. Table 3 shows the sampling of exam specification characteristics for the MCCQE Part II form administered in May 2018. The “Recommended” column specifies the desired number of cases for each exam form per domain and discipline (as shown in Table 1).

Table 3: Sampling of OSCE exam specifications for each test form

Domain	Recommended	May
Counselling	2	3
History	4-5	4
Combined History/PX	(2)(a)	3
Management	2	3
Physical exam	3-4(c)	1
Discipline	Recommended	May
Medicine	3	3
OBGYN	2	2
Pediatrics	2	2
Psychiatry	2	2
Surgery	3	3
Gender(b)		May
As balanced as possible, at least one patient per age group		M = 5
		F = 7
		Either M or F = 0

(a) Combined Hx/Px station may replace a history or physical exam station

(b) Gender of SP may not necessarily be the gender of the actual patient

(c) Physical exams may be combined with management stations or history (Hx/Px) stations

Table 4 shows the frequency (as a percentage) of each of the four reporting domains sampled across the 12 scored stations for the exam form in May 2018. Each domain is sampled a different number of times, with some being measured across all stations and others across a smaller number of stations.

Table 4: Sampling of OSCE subscores¹ for each MCCQE Part II test form

Domain ¹	Target	May
<i>C²LEO</i>	>10%	19%
<i>Data acquisition</i>		35%
<i>Patient/physician interaction</i>		26%
<i>Problem solving/decision making</i>		20%

2.6 EXAM SCORING

Each OSCE station item is assigned to one of four reporting domains: C²LEO, data acquisition, patient/physician interaction and problem solving/decision making. PEs use checklists, oral question items, written items and rating scales² to score the candidate's proficiency on each station. Different PEs evaluate candidates in each of the stations, and one set of station scores per candidate is collected. The ratings provided by each PE are used to calculate all scores.

Table 5 lists each domain along with the types of items that measure each domain.

Table 5: MCCQE Part II reporting domains

DOMAIN	SAMPLING
<i>C²LEO</i>	The C ² LEO content is included in the checklist items, interaction rating scales, oral questions and/or written questions across stations measuring that domain
<i>Data acquisition</i>	The data acquisition content is included in the items in the clinical stations related to history taking and physical examination
<i>Patient/physician interaction</i>	The patient/physician interaction content is included in the rating scale items across most stations
<i>Problem solving/ Decision making</i>	The problem solving and decision-making content is included in items related to the oral questions and from relevant items within the management stations and the written stations

¹ Domain subscores refer to the domains reported to candidates in their Supplemental Feedback Report (SFR).

² Rating scales are six-level Likert items with scores ranging from zero to five. They are typically referred to as patient interaction rating scale items. Most stations include rating scales. The number of items and the specific items vary, depending on the patient problem and task for a station.

In almost all stations, the PE also scores selected rating scale items related to the candidate's interactions with the patient. A complete list of the rating scales used in the MCCQE Part II can be found in Appendix B.

3. CANDIDATE ORIENTATION, EXAM ADMINISTRATION AND TRAINING

3.1 CANDIDATE REGISTRATION AND ELIGIBILITY

To be eligible to take the MCCQE Part II, candidates must pass the MCCQE Part I, their final medical degree diploma must be successfully source verified, and they must have successfully completed a minimum of 12 months of PGT or osteopathic postgraduate training (PGY-2+); or completed 12 months of PGT on, or before the deadline of June 30 for the May exam of the same calendar year or December 31 for the October exam of the same calendar year (PGY-1). The MCC implemented capacity limits to registration for the May and October administrations beginning in fall 2015 due to the demand in particular sessions exceeding the number of exam spaces and the concern that increased demand would affect the quality of the examination administration. Additionally, implementing capacity limits was a way to ensure an adequate sample of candidates to conduct linking of MCCQE Part II total scores. At each administration, PGY-2+ candidates are given priority. If additional spaces are available, PGY-1 candidates are provided the opportunity to apply for the remaining spaces.

For the May 2018 exam, candidates who meet the eligibility requirements stated above were invited to add their names to a MCCQE Part II pre-application list through their physiciansapply.ca account during a specific pre-application period. When the pre-application period ended, candidates were randomly selected from the pre-application list based on available exam capacity. PGY-2+ candidates were given priority on available spaces before the remaining spaces were offered to PGY-1 candidates.

3.2 CANDIDATE ORIENTATION

The MCC provides candidates with detailed information about the MCCQE Part II on its website. Topics include what to expect on examination day, scoring, results, application information, and test accommodations. Candidates also have access to an online orientation presentation. In

addition, they must participate in a mandatory orientation given by senior site staff on each exam day before the exam begins. These sessions provide candidates with:

- Information on the personal belongings that they can and cannot bring to the exam
- Information on how they may use their booklet and their bar code labels
- An overview of the length and number of stations and how they will rotate through the stations
- Information on the timing of the stations and how the signal system functions
- Guidance on how to interpret the candidate instructions
- Information on how they will be assessed by the PEs and when a PE may intervene
- Information on available medical equipment
- A description of how they should interact with the SPs
- An overview of the short-answer write-ins and the multiple-choice written stations
- Instructions on exam security and how to ask for assistance
- Information about the MCC's conflict of interest policy, confidentiality, sequestration, expected candidate behaviour, and how to report concerns on exam day and to the MCC after the exam

3.3 EXAM ADMINISTRATION

In May 2018, 2,573 candidates participated in the MCCQE Part II administration. The exam was administered in English in Calgary, Edmonton, Halifax, Hamilton, Kingston, London, Montreal, Ottawa, Saskatoon, St. John's, Sudbury, Toronto, Vancouver, Victoria and Winnipeg. The exam was also offered in French in Montreal, Quebec City and Sherbrooke.

3.4 EXAM ADMINISTRATIVE STAFF

Each partner site is responsible for hiring and supervising administrative staff. They work with the MCC to safeguard exam materials and are accountable for the overall coordination of the exam team's activities.

MCC personnel oversees site staff on exam days across the country in person, by telephone, and via electronic communication and works a hotline on exam days.

Every two years, the MCC hosts a face-to-face meeting for all MCCQE Part II site administrators to review administrative aspects of the examination, troubleshoot site-specific challenges such as examiner recruitment and enhance collaboration between sites.

3.5 STANDARDIZED PATIENT TRAINING

Each site is responsible for hiring and supervising the SP trainers who oversee the SPs and assure the quality of their standardized performances on exam day(s). SPs are trained at each site using MCCQE Part II training materials provided by the MCC. Training support is provided centrally by MCC staff, primarily by the MCCQE Part II training officer.

Every two years, the MCC hosts a face-to-face meeting for all MCCQE Part II SP trainers to review SP training aspects of the exam, troubleshoot site-specific challenges such as SP recruitment and to enhance collaboration between sites and the MCC.

3.6 DEPUTY REGISTRARS

The Deputy Registrar (DR), on behalf of the Dean, is responsible for the administrative and financial operations of the exam centre and the local administration of the MCCQE Part II. The DR is also responsible for:

- Ensuring that an office is established for the team, adequate office space and furnishings are available, including computer equipment, phones and copiers, and that security precautions are taken
- Selecting and supervising centre personnel, especially the exam administrator as well as hiring senior site staff, a task usually performed in consultation with the manager of the MCCQE Part II
- Providing leadership and support to senior site staff for planning and teamwork
- With assistance of the Chief Examiners (CE), recruiting physicians to serve as examiners, assigning them to their stations and conducting pre-exam orientation sessions for both first-time and experienced examiners
- Delivering orientation to CEs
- Assisting with dry runs for the SPs for Quality Assurance (QA), a role usually shared with the CEEs ensuring the administration runs smoothly, especially in the following areas:
 - Checking props and room setup
 - Supervising and supporting staff throughout the day
 - Ensuring the examination centre is securely closed at the end of the exam day
- Reviewing all incident reports and ensuring appropriate action is taken and documented
 - Incident reports are completed to report the details of unusual events or breaches in procedures to the MCC; these reports may be completed by candidate(s), PEs, SPs, SP trainers, the DR, the CE, the Site Administrator or site staff
- Liaising with MCC before, during and after the exam

- Overseeing the preparation of a financial report, and interim financial report, if necessary, and the DR's report for the exam session
- If requested, reviewing and commenting on training materials, cases and other documentation before deadlines

3.7 CHIEF EXAMINERS

The CE's role depends on the size of the centre and how the DR assigns responsibilities. Generally, the CE should assist with SP dry runs, participate in some of the SP training sessions and assist in examiner recruitment and staff training as needed. All SPs take part in a dry run prior to exam day with the CE playing the candidate's role to ensure they are ready to perform their roles. An SP who is deemed as not being ready is re-trained or replaced.

3.8 PHYSICIAN EXAMINER RECRUITMENT AND TRAINING

Table 6 presents the requirements used to recruit PEs.

Table 6: PE recruitment requirements

REQUIREMENTS FOR PE RECRUITMENT	
1.	PEs must be registered and in good standing with a Medical Regulatory Authority (MRA) in Canada.
2.	PEs can be semi-retired or retired, but they must have an active licence with an MRA in Canada.
3.	PEs must agree to adhere to the MCC's Code of Business Conduct (the Code): <ul style="list-style-type: none"> • The Code addresses, but is not limited to, the application of the Code, legal and ethical standards, standards of conduct and compliance with the Code. • Any breach of the Code could lead to legal and/or disciplinary measures against the responsible persons, including without limitation their removal from office, their expulsion from committees or the termination of the applicable contract.
4.	PEs must have the ability and stamina to complete the task (e.g., uncorrected hearing loss is a serious handicap).
5.	PEs must be practising or have practised in Canada within the last five years.
6.	PEs must have the Licentiate of the Medical Council of Canada (LMCC) and must provide their LMCC registration number. PEs must be two years post-LMCC.

7.	PEs must have at least two years of independent practice in Canada. (Exception: for residents, see number 9.)
8.	PEs do not need to be faculty members if all other criteria are met AND they meet the following criterion: <ul style="list-style-type: none"> • They must have recent experience supervising clerks and residents (within the last two years) (e.g., a physician who has worked for a university as an OSCE examiner or for other similar OSCEs).
9.	PEs can be residents, but they must be PGY5 level or higher, or they can be fellows, OR they can have CCFP certification. <ul style="list-style-type: none"> • Note: Residents must not exceed 20 per cent of the examiner cohort assigned to a given session.
10.	Physicians who do not have their LMCC will be accepted as examiners under the following conditions: <ul style="list-style-type: none"> • Non-licentiate PEs must be faculty members (e.g., faculty lecturer, assistant professor, associate professor or professor). <p>AND</p> <ul style="list-style-type: none"> • Non-licentiate PEs must be certified by one of the following organizations and must provide their certification number: <ul style="list-style-type: none"> ◦ Royal College of Physicians and Surgeons of Canada (RCPSC) ◦ Collège des médecins du Québec (CMQ) ◦ College of Family Physicians of Canada (CFPC) <p>AND</p> <ul style="list-style-type: none"> • Non-licentiate PEs must sign a waiver indicating that they have no intention of taking the MCCQE Part II exam.
11.	Note: Non-licentiate PEs must not exceed 50 per cent of the examiner cohort assigned to a given session.
12.	Any potential PEs who do not meet the above guidelines must be pre-approved by the MCCQE Part II Exam Manager. Documentation of qualifications may be requested.

The MCC provides an exam day PE orientation given by the sites' CE, DR and/or senior site staff. Before every exam, all PEs must also participate in an online PE orientation that provides information on enhancing the standardization of PE scoring, including a practice scoring session and a guided discussion.

4. MCCQE PART II SCORING

This section describes quality assurance and quality control procedures relating to the scoring of candidates on the MCCQE Part II, what scores are reported and how they are calculated.

4.1 STANDARD QUALITY ASSURANCE AND QUALITY CONTROL PROCEDURES

To ensure the accuracy and integrity of candidates' exam day records, QA steps are performed as outlined below.

PEs complete a score sheet for each candidate seen in their OSCE stations. These forms are scanned at each exam site and transmitted securely to the MCC where MCC staff import the forms into Cardiff TeleForm[®] software where they are reviewed. Scanning anomalies are identified (for example, a non-scannable candidate barcode label, a PE's pencil marks that are too faint, missing sheets due to candidates that do not complete the exam on both days, etc.) and the requisite corrections are made to the electronic data records. Data is imported electronically into a scoring application to generate a list of all candidates whose total score is close to (approximately +/-3 per cent) the pass score. We call this group the selected candidate group. Paper copies of the score sheets for this selected candidate group are visually reviewed. Some examples of checks include:

- Number of selections per question for extended match questions where candidates can bubble a number of correct answers
- Notes by PEs for oral or SP questions
- Confirmation of missing data for oral questions or rating scales
- Flags for lapses in patient safety or professional behaviours
- Confirmation of scores and bubbles for written items
- Verification of raw score points

Any differences are corrected in the electronic data files to reflect the paper score sheets. There are typically 100 to 120 corrections made on each exam. The updated electronic files are then re-imported into the scoring application that is used to create the scale scores for all candidates. All scores are calculated independently in parallel using the *Statistical Analysis System* (SAS[®]) and compared to the results from the scoring application. All values must match before results are released to candidates.

4.2 EXAM RESULT APPROVAL

The results for each administration of the MCCQE Part II are reviewed by the CEC. The CEC approves the release of results after each administration, including special cases (described below). Once the CEC has approved the results, they are imported and released to candidates.

Exam results that meet the following criteria are automatically approved by the CEC: (1) station and item level p-values between 0.10 and 0.90, (2) Station Total Correlation (STC) greater than 0.30 and Item Total Correlation (ITC) greater than 0.05, and (3) decision consistency and accuracy values greater than 0.80. If a p-value, STC or ITC is not within the approved values, the content for that item is reviewed by subject matter experts to ensure that the content is still appropriate and defensible. Automatically approved exam results are reported to candidates once all the QA processes are completed. Special cases are not automatically approved.

When an incident occurs that may impact a candidate's performance on exam day, it is presented to the CEC as a special case. The CEC determines the severity of the impact and decides if any change should be made to the candidate's exam results. Depending on the nature of the incident, the CEC may decide to remove one or more items from a candidate's exam, award the candidate a "No Standing", or a "Denied Standing".

A "No Standing" indicates that procedural irregularities in the examination process may have materially affected the performance of the candidate and/or may have prevented a reliable assessment of the candidate's knowledge and abilities. A 'No Standing' does not count towards a candidate's number of attempts.

A "Denied Standing" indicates that a candidate has been found to have committed an infraction related to the MCC's examination process and/or breached the confidentiality of the examination. A 'Denied Standing' counts as an attempt towards a candidate's total number of attempts. Additionally, candidates that are awarded a "Denied Standing" may be denied entry to one or more future examinations of the MCC.

4.3 EXAM RESULT REPORTING

Approximately seven weeks after the last day of the exam session, the MCC issues a Statement of Results (SOR) and Supplemental Feedback Report (SFR) to each candidate through their physiciansapply.ca account (samples from May 2018 administration are shown in Appendix C). The SOR includes the candidate's final result and total score, as well as the pass score. Additional information about a candidate's domain subscores and comparative information is provided on the SFR in graphic form.

The total score is reported on a standard-score scale ranging from 50 to 950. In contrast, the score profile in the SFR example displays a candidate's domain subscores in terms of a percentage. As a result, total scores cannot be compared to domain subscores in the SFR as they are reported on different scales. Additionally, it is important to note that, because subscores have fewer items, there is less measurement precision than total scores. Subscores are provided to individual candidates for feedback only and are not meant to be used by organizations for selection.

The following sections outline the steps in creating the results reported to candidates.

4.3.1 Scale score

Deriving the scale score for the MCCQE Part II involves three steps.

Step 1: Calculate total scores

The first step in obtaining a scale score is to calculate the total score for each candidate. To do so, a station score is calculated for each station using the following formula:

$$\text{station score} = \frac{\sum_i \text{score}_i}{\sum_i \text{max_score}_i} * 100$$

where the numerator is the sum of each candidate's scores on each item *i* for that station and the denominator is the sum of the maximum score for each item for that station. For example, a station with several checklist items, an oral question, and several rating scales could result in the following score:

$$\text{station score} = \frac{\text{sum of items}=60}{\text{sum of max. score of items}=67} * 100=89.55$$

The station scores are then used to calculate the total score for each candidate using the following formula:

$$\text{total score} = (\text{sum of 12 station scores})/12$$

Since station scores are based on the sum of the candidate's items for that station, missing data needs to be taken into account so that it does not negatively impact a candidate's score. Missing data occurs when the PE does not provide a rating for some of the oral questions or rating scales for a given candidate on the scannable score sheet. When oral questions or rating scales are not scored by the PE, the station score is based on the items that are provided by the PE. Extending the example above, a station that has one missing rating scale item with a maximum score of five would result in the following score:

$$\text{station score} = \frac{\text{sum of candidate items}=60}{\text{sum of max. score of items}=62} * 100=96.77$$

The station score would have been 89 per cent if the missing rating scale was treated as zero and the adjustment not applied. However, to be fair to the candidate, we exclude the missing rating scale from the calculation of the station score.

Step 2: Link MCCQE Part II scores to base form – spring 2015

For each MCCQE Part II administration, a different test form is used to measure the exam specifications (as described in the exam specifications section). The MCC staff and the MCCQE Part II Test Committee work in collaboration to select eight 10-minute OSCE stations and four couplet OSCE stations that best meet exam specifications. While good for security reasons, one possible adverse effect of having different exam forms with different stations across administrations is that one test form may be more difficult than the other. However, the process of linking total scores statistically takes into account differences in test form difficulty and adjusts total scores so that all scores are on the same scale. Linking is also a way of applying the same pass score to candidates who take different test forms.

The next step in obtaining scale scores for all candidates is to link scores through the common stations to the base form. For the May 2018 MCCQE Part II test form, linking occurred to place the candidate's scores on the same scale as the base form from the spring 2015 MCCQE Part II administration.

One method of linking test forms is to have a subset of the content appear identically across test forms; this is called a non-equivalent anchor test or NEAT design. The subset of content that is presented identically is called an anchor set. The rule of thumb for determining the number of items in an anchor set for a multiple-choice exam is 20 per cent of the total test or 20 items, whichever is greater, to ensure that the anchor set is representative of the total test in terms of content and difficulty. Since the MCCQE Part II is an OSCE with a small number of stations (less than 20), we use a 30 per cent rule. The anchor set is used to statistically estimate the overall ability of candidates, taking each test form and the difficulty of the test forms into account.

For the MCCQE Part II May 2018 test form, the anchor set was based on three stations. A reference group of Canadian medical graduate and Canadian postgraduate first-time test takers was used for all linking calculations. The linking calculations from this reference group is applied to all candidates to calculate each candidate's linked score. This linked score is then transformed as described in step 3 below. In all linking steps, the Levine's observed score method was employed. Full details of the method can be found in *Testing*

equating, scaling, and linking: Methods and Practice (2nd Edition) by Kolen and Brennan (2004).

Step 3: Transform linked scores to scale scores

Once the first two steps are complete, the linked scores are transformed to scale scores ranging from 50 to 950 for reporting purposes. Using the spring 2015 MCCQE Part II results from all candidates, the new scale was established to have a mean of 500 and a Standard Deviation (SD) of 100. This final transformation ensures that any differences in candidate scale scores, or mean scale scores and SD on subsequent test forms can be directly compared to the spring 2015 MCCQE Part II results. For example, a candidate score or mean cohort increase to 600 or decrease to 400 would indicate the general performance of the candidate or group of candidates who took the subsequent test forms was one SD higher and lower, respectively.

The final transformation formula for all test forms is as follows:

$$\text{ScaleScore}_X = (\text{slope})(\text{LinkedScore}_X) + (\text{intercept})$$

ScaleScore_X is the linear function used to calculate the scale score for candidate X, where the slope is equal to 15.08, the intercept is equal to -459.86, and the LinkedScore_X is the linked score for candidate X.

Finally, all scale scores are rounded to a whole number between 50 and 950. The reported scale scores as seen by candidates are these rounded values. For example, a candidate with a linked score of 81.25 would have a scale score of 765:

$$\text{ScaleScore}_X = (15.08) * (81.25) + (-459.86) = 765.39 \text{ rounded to } 765$$

Another candidate with a linked score of 42.51 would result in a scale score of 181:

$$\text{ScaleScore}_X = (15.08) * (42.51) + (-459.86) = 181.19 \text{ rounded to } 181$$

4.3.2 Pass/fail status

The MCC completed a rigorous standard-setting exercise using the borderline group method and a panel of 20 physicians from across the country that represented faculties of medicine, different specialties, and had differing years of experience supervising residents. Full details of the standard-setting exercise can be found in the [*Technical Report on the Standard-Setting Exercise for the Medical Council of Canada Qualifying Examination Part II*](#) (Medical Council of Canada, 2015). The pass score of 509 was recommended by the panel of physicians for spring 2015 and approved by the CEC in June 2015. This pass score was

used to assign a pass/fail status to each candidate of the May 2018 administration. A score of 509 or greater is deemed a pass.

4.3.3 Domain subscores

Domain subscore calculations are used to create the score profile in the candidates' SFRs. For each domain subscore, the associated items are converted to a percentage ranging from 0 to 100, where the total number of score points obtained by a candidate is divided by the maximum score points per domain, multiplied by 100.

For example, if a candidate received scores of five, seven, eight and one on a domain with associated maximum scores of 10, 10, 9 and 1, the total number of score points obtained by the candidate is 21; the maximum number of score points for this domain is 30. The domain score is $21/30 \times 100$ or 70.0. For the MCCQE Part II, there are four domain subscores that are presented to candidates in their SFRs: C²LEO, data acquisition, patient/physician interaction and Problem Solving and Decision Making (PSDM).

As a reminder, domain subscores should not be compared to total scores as they are reported on different scales and because they have fewer items than total scores, they have less measurement precision. Domain subscores are intended to provide general feedback to candidates on their relative strengths and weaknesses on their performance on the MCCQE Part II.

5. PSYCHOMETRIC RESULTS

The data used for the aggregate analyses are the results approved by the CEC in June 2018 for the May 2018 administration. In total, 4,573 candidates participated in the MCCQE Part II administered in May 2018.

The number of candidates for the May 2018 administration is depicted in Table 7. It shows the number of candidates for the May 2018 administration by PGT group (PGY-1 and PGY-2+) by candidate group (for example, Canadian Medical Graduates (CMG), first-time test takers (1st), etc.), gender and examination language. The main reference group for this examination includes CMGs and the Canadian postgraduate first-time test taker candidate group. Candidate groups shown in Table 7 are as follows:

- CMG, Canadian postgraduate, first-time test takers (CMG-CPG 1st)
- CMG, Canadian postgraduate, repeat test takers (CMG-CPG repeaters)
- CMG, international postgraduate, first-time test takers (CMG-IPG 1st)

- CMG, international postgraduate, repeat test takers (CMG-IPG repeaters)
- International medical graduate (IMG), international postgraduate, first-time test takers (IMG-IPG 1st)
- IMG, international postgraduate, repeat test takers (IMG-IPG repeaters)
- IMG, Canadian postgraduate, first-time test takers (IMG-CPG 1st)
- IMG, Canadian postgraduate, repeat test takers (IMG-CPG repeaters)

Table 7: Number and percentage of candidates for the MCCQE Part II by PGT group

PGT Group	Subgroups	May 2018 ^(a)	
PGY-1	Candidate group	N	%
	CMG-CPG 1st	572	87.2
	CMG-CPG repeaters	0	0.0
	CMG-IPG 1st	0	0.0
	CMG-IPG repeaters	0	0.0
	IMG-IPG 1st	2	0.3
	IMG-IPG repeaters	0	0.0
	IMG-CPG 1st	78	11.9
	IMG-CPG repeaters	4	0.6
	Gender	N	%
	Female	352	53.7
	Male	304	46.3
	Language	N	%
	English	520	79.3
	French	136	20.7
	Total	656	100.0
	PGY-2+	Candidate group	N
CMG-CPG 1st		874	45.6
CMG-CPG repeaters		76	4.0
CMG-IPG 1st		5	0.3
CMG-IPG repeaters		0	0.0
IMG-IPG 1st		496	25.9
IMG-IPG repeaters		236	12.3
IMG-CPG 1st		169	8.8
IMG-CPG repeaters		61	3.2
Gender		N	%
Female		867	45.2
Male		1050	54.8
Language		N	%
English		1606	83.8
French		311	16.2
Total		1917	100.0

(a) One “No Standing” candidate and five “Denied Standing” candidates are not included in the remaining analyses. One additional candidate is not included in the remaining statistical analyses (e.g., mean etc.) due to missing more than one station, but this candidate is included in the pass rate and N.

In the sections below, summary statistics for scale scores and pass rates are provided, as well as estimates of reliability for both scores and classification decisions. Finally, a summary of station scores and domain subscore profiles are provided.

5.1 SCALE SCORES

Scale score summary statistics and pass rates for the May 2018 MCCQE Part II are presented in Table 8 for the reference group (CMG-CPG 1st) and total group.

Table 8: Summary statistics of scale scores by PGT group

Exam session	Candidate group	PGT group	N	Min.	Max.	Mean	Median	SD	Pass rate
May 2018 ^(a)	Total	PGY-1	656	454	764	627.4	628	54.1	97.6
		PGY-2+	1911	296	780	588.9	594	81.6	83.5
	CMG-CPG 1st	PGY-1	572	464	764	631.7	633	52.0	98.6
		PGY-2+	874	445	780	638.4	641	59.3	97.9

One candidate is not included in the remaining statistical analyses (e.g., mean, etc.) due to missing more than one station, but this candidate is included in the pass rate and N.

Figure 1 displays the distribution of scale scores for the MCCQE Part II candidates for May 2018, by PGT Group. Figure 1 shows that a smaller proportion of candidates were in the PGY-1 group than the PGY-2+ group, and the candidates in PGY-1 group were slightly higher performing than the PGY-2+ group.

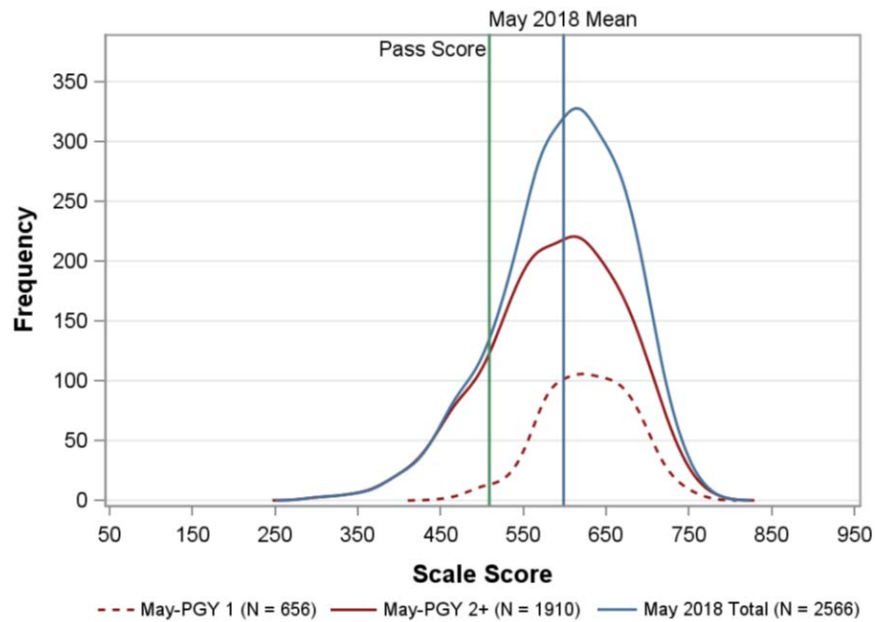


Figure 1: Scale score distribution for May 2018 for total candidates and by PGT group

Figure 2 displays the distribution of scale scores for May 2018 by PGT group for CMG-CPG 1st candidates. Figure 2 shows that a smaller proportion of candidates were in the PGY-1 group than the PGY-2+ group, and the candidates in PGY-1 group were slightly lower performing than the PGY-2+ group.

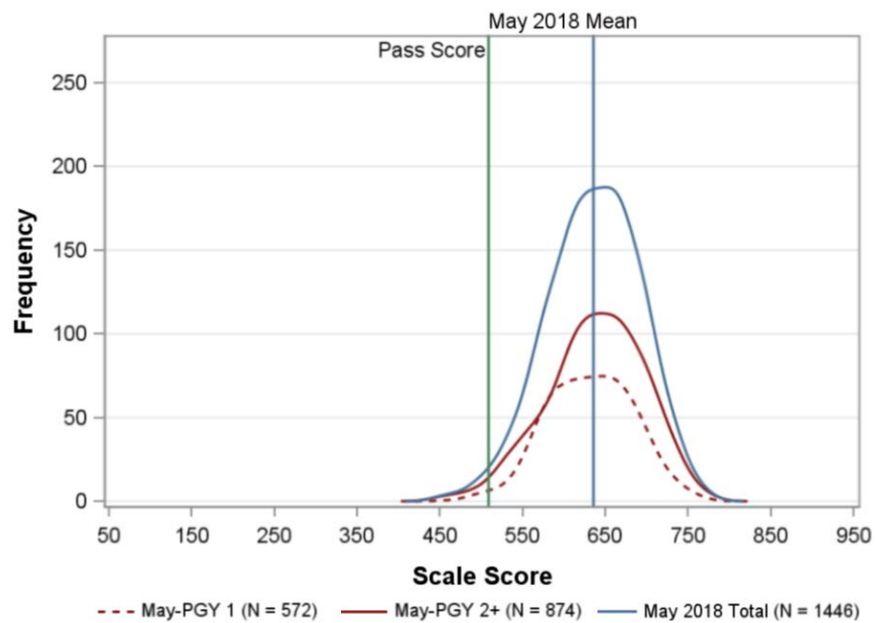


Figure 2: Scale score distribution for May 2018 for CMG-CPG first-time test takers, by PGT group

5.2 ESTIMATES OF RELIABILITY AND CLASSIFICATION DECISIONS

5.2.1 Cronbach's Alpha

Cronbach's alpha was used to estimate score reliability for the MCCQE Part II test form. A score reliability estimate indicates the desired consistency (or reproducibility) of examination scores across replications of measurement (Crocker & Algina, 1986; Haertel, 2006). Scores that are highly reliable are accurate, reproducible and consistent from one testing occasion to another. In other words, if the testing process was repeated with a group of test takers, essentially the same results would be obtained. The reliability estimate is further described in Educational Measurement by Haertel in chapter 2, section 2.4.4 (Haertel, 2006). The formula for Cronbach's alpha is:

$$\alpha \rho_{XX'} = \frac{n}{n-1} \left(1 - \frac{\sum \sigma_{X_i}^2}{\sigma_X^2} \right)$$

where n is the number of stations, $\sigma_{X_i}^2$ is the variance of station i scores and σ_X^2 is the variance of total scores (Haertel, 2006, p. 74). As a general rule, a reliability estimate is desired to be greater than 0.80 on an OSCE. The reliability estimate, in conjunction with the total exam Standard Error of Measurement (SEM), provides further evidence of the reliability of the candidate's total score.

5.2.2 Standard error of measurement

The SEM provides a value within a certain confidence range (for example, +/- 1 SEM and +/- 2 SEMs represent a 68 per cent or 95 per cent range, respectively) that a candidate's observed score is expected to range if the candidate was retested over repeated exams that meet the same exam specifications. The smaller the SEM, the less measurement error in a candidate's score. The SEM is calculated as follows:

$$SEM = SD \sqrt{1 - \alpha \rho_{XX'}}$$

where SD is defined as the standard deviation of total scores (square root of the variance) and $\alpha \rho_{XX'}$ is defined as the reliability estimate as shown above.

5.2.3 Decision consistency and decision accuracy

Estimates indicating the consistency and accuracy of pass/fail decisions are important in providing validity and reliability evidence for candidate scores on one test form with possible equivalent test forms. To this end, the MCCQE Part II uses the Livingston and

Lewis (1995) procedure, where decision consistency is an estimate of the agreement between classifications on potential parallel test forms, and decision accuracy is the estimate of agreement between the observed classifications of candidates and those based on their estimated true score (observed score \pm measurement error). Ideally, both values should be high, such as 0.80 and above, supporting a reliable and valid pass/fail standing. A value of 0.80 indicates that accuracy or consistency of the decision is being met for at least 80 per cent of the candidates.

Table 9 shows the reliability estimates, the SEM and the decision consistency, and decision accuracy estimates along with the associated false positives and false negative rates for the May 2018 test form. The estimated false positive rate indicates the expected proportion of candidates who pass based on their observed scores but who should fail based on their estimated true ability. The estimated false negative rate indicates the expected proportion of candidates who fail based on their observed scores but who should pass based on their estimated true ability.

Table 9: Reliability estimates, SEM, decision consistency and accuracy for May 2018

	May 2018
Reliability estimate	0.77
SEM (score scale)	37.05
Decision consistency	0.90
False positive	0.05
False negative	0.05
Decision accuracy	0.93
False positive	0.02
False negative	0.04

It should be noted that reliability is impacted both by the amount of variability in scores amongst candidates taking a particular test form and the number of items or stations included in any given exam. It is more difficult to obtain reliability estimates above 0.80, given the restricted number of stations that can be administered in any OSCE form. The decision consistency and accuracy values should be above 0.8 for OSCEs and for May 2018 these values were well above 0.8.

5.3 OSCE STATION STATISTICS

Summary statistics for each of the OSCE stations are provided in Table 10. The percentage of missing data, average station score or p-value, SD of stations scores and STCs are presented. Please refer to 4.3.1 for calculation of station scores.

P-values are the average station scores that candidates achieved on each of the stations. In general, *p*-values indicate station difficulty and range between 0 and 1. Station *p*-values that are low (<0.20) indicate a difficult station; those that are high (>0.90) indicate an easy station. *P*-values are population dependent. That is, comparisons of *p*-values across different samples of candidates do not take into account potential differences in overall candidate ability. As such, *p*-values should not be overinterpreted or used as the only indicator of difficulty. Rather, *p*-values provide a general sense of the range of difficulty of stations on a particular test form.

SD indicate the general variability of scores on any given station. STCs are indicators of discrimination between low- and high-ability candidates for a given station. A low positive or negative STC (i.e., -0.20 to <0.30) indicates that there is a weak to negative relationship between the station score and the overall exam score. Along with the *p*-values, this information is useful in flagging stations that should be reviewed by content experts and possibly removed from scoring. A moderate to high STC (>0.30) indicates that high-ability candidates are performing well on a given OSCE station. Stations with STCs that are below 0.30, as well as negative values, are flagged for content review. Flagged and reviewed stations may still be included on an exam when the content is deemed relevant, important and has been verified to be correct.

Table 10: Summary statistics for OSCE stations for May 2018

Station	May 2018			
	% missing	p-value	SD	STC
1	0.02	0.68	0.11	0.36
2	0.06	0.77	0.14	0.37
3	0.01	0.70	0.12	0.42
4	0.03	0.70	0.12	0.41
5	0.06	0.72	0.11	0.37
6	0.02	0.67	0.11	0.35
7	0.01	0.67	0.14	0.46
8	0.03	0.68	0.13	0.39
9	0.05	0.68	0.15	0.46
10	0.02	0.72	0.11	0.41
11	0.01	0.65	0.13	0.44
12	0.01	0.67	0.13	0.43
Min.	0.01	0.65	0.11	0.35
Max.	0.06	0.77	0.15	0.46
Mean	0.03	0.69	0.12	0.41
SD	0.02	0.03	0.01	0.04

Table 10 shows the mean p -values for each administration. There were no stations flagged as being too difficult (p -value <0.30) or too easy (p -value >0.90). Stations with items with an ITC <0.05 were reviewed for content appropriateness. One item score key was adjusted after subject matter experts indicated a change was required. This adjusted score key was used for all candidates.

5.4 EXAMINER ANALYSES

Examiner analyses are conducted routinely for each of the 10-minute stations for each PE. For the couplet stations, the examiner analyses are conducted only for the patient interaction component. For the examiner analyses, the following steps are followed:

Step One

For each PE and station/component the PE scored, the average across the candidates' station/component scores is calculated. This average is the PE average for that station or component. Then the average of the PE averages is calculated along with the SD. PEs that scored fewer than 10 candidates on a station are excluded from these analyses as they have observed too few candidates to be compared to other PEs. PEs are flagged as being a "Dove" if their average station or component score is more than three times the station/component SD from the station/component average. PEs are flagged as being a "Hawk" if their station or component score is less than three times the station/component SD from the station/component average. For example, if the average across PE averages was 72.5 and the SD across PEs was 6.5 and a PE had an average of 50.7 [difference of 21.8, which is greater than 3SDs ($6.5 \times 3 = 19.5$)] then he/she is flagged as a "Hawk".

Step Two

In step two, for each PE flagged in step one, the station distribution (histogram) for the PE is compared to the distribution of station scores from all other PEs across the country. This is a visual check to evaluate whether the PE is providing a range of scores that looks somewhat normally distributed (not providing all high or low scores). If a PE's distribution looks reasonable, they are no longer flagged at this step as being either a "Dove" or "Hawk".

Step Three

In step three, for each PE flagged in step one and two, the reported score distribution (histogram) for the cohort they scored is compared to the distribution of scale scores from the rest of the candidates across the country. This is a check that the cohort based on all 12 examiners who scored each of their complete exam is higher or lower than the values across the country. In this step, we evaluate if a cohort may be higher or lower in ability that may explain a "Dove" or "Hawk"

flag in step one. In addition, the average scale scores and pass rate for this PE's cohort are compared across the country. In this step, we are evaluating whether the cohort of candidates being evaluated may be a weaker or stronger in ability based on all 12 stations. For example, a PE may be flagged as being a "Hawk" in steps one and two, but their scale scores based on all 12 stations may have substantially lower scale scores, indicating a weaker cohort, and thus would not be flagged as a "Hawk" at step three.

One PE for May 2018 was flagged across all three steps as a "Hawk". This PE was provided a feedback letter on their performance as an examiner, along with the option of a follow-up conversation with the MCCQE Part II Manager.

5.5 DOMAIN SUBSCORE PROFILES

The purpose of the domain subscore profile is to provide diagnostic feedback to candidates by highlighting their relative strengths and weaknesses in four areas. A domain subscore profile is presented in the form of a graph to each candidate in the SFR. The graph shows the domain subscore for each of the four domains and the SEM around the domain subscore. The calculation of the domain subscores for each candidate is outlined in Section 4.3.3.

This section provides domain subscore profiles for all candidates for the May 2018 MCCQE Part II results. The range of domain subscores for the May 2018 test form is shown graphically in Figure 3. The boxes for each domain indicate the range for 50 per cent of the candidates' domain subscores. The vertical line represents the median or 50th percentile domain subscore. The remaining 50 per cent of domain subscores are shown to the right or left of the box as a line (25 per cent to the right and 25 per cent to the left). The mean domain subscores are indicated by the diamond. The note under each figure indicates the naming convention for each domain.

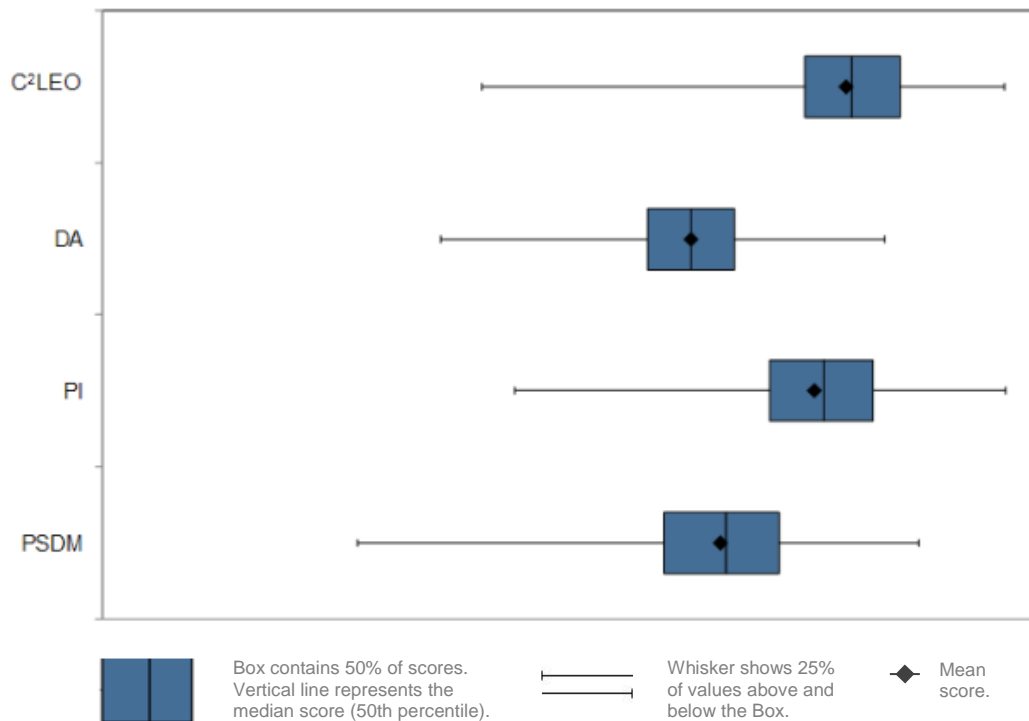


Figure 3: Domain subscore for May2018 MCCQE Part II candidates

Note: C²LEO = Considerations for Cultural Communication, Legal, Ethical, and Organizational aspects of the Practice of Medicine, DA = Data Acquisition, PI = Patient / Physician Interaction, PSDM = Problem Solving and Decision Making

5.6 HISTORICAL PASS RATES

Historical pass rates since harmonization with the College of Family Physicians of Canada (CFPC) in 2013 are presented in this section. Table 11 shows the pass rates from 2013 to May 2018 for four subgroups and the total group of candidates that took the MCCQE Part II or the equivalent CFPC Clinical Skills Exam (CSE) test forms. The four subgroups are CMG-CPG 1st, CMG-CPG Repeat, Other 1st and Other Repeat, where Other is a mix of undergraduate medical education or postgraduate training outside of Canada. Tables 12 and 13 show the pass rates from 2013 to May 2018 for the same four subgroups and total group by spring and fall test administration of each year (excluding October 2018).

Table 11: 2013 – May 2018 pass rates by total reference group

Year	Candidate group	N total	N pass	Pass rate
May 2018	Total	2567	2235	87.1
	CMG-CPG 1st	1446	1420	98.2
	CMG-CPG Repeat	76	69	90.8
	Other 1st	746	565	75.7
	Other Repeat	299	181	60.5
2017	Total	4981	4340	87.1
	CMG-CPG 1st	2871	2791	97.2
	CMG-CPG Repeat	265	241	90.9
	Other 1st	1170	870	74.4
	Other Repeat	675	438	64.9
2016	Total	4919	3906	79.4
	CMG-CPG 1st	2969	2719	91.6
	CMG-CPG Repeat	282	226	80.1
	Other 1st	1020	638	62.5
	Other Repeat	648	323	49.8
2015	Total	4806	3611	75.1
	CMG-CPG 1st	2792	2574	92.2
	CMG-CPG Repeat	213	145	68.1
	Other 1st	1162	651	56.0
	Other Repeat	639	241	37.7
2014	Total	4472	3467	77.5
	CMG-CPG 1st	2554	2388	93.5
	CMG-CPG Repeat	149	111	74.5
	Other 1st	1118	694	62.1
	Other Repeat	651	274	42.1
2013	Total	4204	3320	79.0
	CMG-CPG 1st	2476	2361	95.4
	CMG-CPG Repeat	128	102	79.7
	Other 1st	972	578	59.5
	Other Repeat	628	279	44.4

**Table 12: Spring 2013 – May 2018
pass rates by total reference group**

Year	Candidate group	N total	N pass	Pass rate
2018	Total	2567	2235	87.1
	CMG-CPG 1st	1446	1420	98.2
	CMG-CPG Repeat	76	69	90.8
	Other 1st	746	565	75.7
	Other Repeat	299	181	60.5
2017	Total	2423	2050	84.6
	CMG-CPG 1st	1163	1139	97.9
	CMG-CPG Repeat	203	188	92.6
	Other 1st	652	466	71.5
	Other Repeat	405	257	63.5
2016	Total	2363	1771	74.9
	CMG-CPG 1st	1248	1120	89.7
	CMG-CPG Repeat	179	144	80.4
	Other 1st	549	317	57.7
	Other Repeat	387	190	49.1
2015	Total	2237	1656	74.0
	CMG-CPG 1st	1158	1109	95.8
	CMG-CPG Repeat	126	88	69.8
	Other 1st	620	350	56.5
	Other Repeat	333	109	32.7
2014	Total	2064	1611	78.1
	CMG-CPG 1st	1082	1047	96.8
	CMG-CPG Repeat	67	46	68.7
	Other 1st	603	386	64.0
	Other Repeat	312	132	42.3
2013	Total	1826	1430	78.3
	CMG-CPG 1st	995	958	96.3
	CMG-CPG Repeat	69	52	75.4
	Other 1st	446	273	61.2
	Other Repeat	316	147	46.5

Table 13: Fall 2013 – 2017 pass rates by total reference group

Year	Candidate group	N total	N pass	Pass rate
2017	Total	2558	2290	89.5
	CMG-CPG 1st	1708	1652	96.7
	CMG-CPG Repeat	62	53	85.5
	Other 1st	518	404	78.0
	Other Repeat	270	181	67.0
2016	Total	2556	2135	83.5
	CMG-CPG 1st	1721	1599	92.9
	CMG-CPG Repeat	103	82	79.6
	Other 1st	471	321	68.2
	Other Repeat	261	133	51.0
2015	Total	2569	1955	76.1
	CMG-CPG 1st	1634	1465	89.7
	CMG-CPG Repeat	87	57	65.5
	Other 1st	542	301	55.5
	Other Repeat	306	132	43.1
2014	Total	2408	1856	77.1
	CMG-CPG 1st	1472	1341	91.1
	CMG-CPG Repeat	82	65	79.3
	Other 1st	515	308	59.8
	Other Repeat	339	142	41.9
2013	Total	2378	1890	79.5
	CMG-CPG 1st	1481	1403	94.7
	CMG-CPG Repeat	59	50	84.7
	Other 1st	526	305	58.0
	Other Repeat	312	132	42.3

6. REFERENCES

- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Sage Publications Inc.: Thousand Oaks, CA.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart & Winston, Inc.: Orlando, FL.
- Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65-110). Westport, CT: Praeger Publishers.
- Kane, M.T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp.17-64). Westport, CT: Praeger Publishers.
- Kane, M.T. (2013). Validating the interpretations and uses of test scores. *Journal of educational measurement*, 50(1), 1-73.
- Kolen, M.J., & Brennan, R.L., (2004). *Test equating, scaling, and linking: Methods and practice* (2nd ed.). New York, NY: Springer Science Business Meeting.
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of educational measurement*, 32(2), 179-197.
doi: 10.1111/j.1745-3984.1995.tb00462.x
- Medical Council of Canada. (2014). *Blueprint project: Qualifying examinations blueprint and content specifications*. MCC: Ottawa, Canada. Retrieved from:
mcc.ca/wp-content/uploads/Blueprint-Report.pdf
- Medical Council of Canada. (2015). *Technical report on the standard setting exercise for the Medical Council of Canada Qualifying Examination Part II*. MCC: Ottawa, Canada. Retrieved from: mcc.ca/wp-content/uploads/MCCQE-Part-II-Standard-Setting-Report-July2015.pdf

APPENDIX A: QUALITY CONTROL – MCCQE PART II RESULTS

Pre-examination

1.	Design sheets, set up data exports and activate sheets for all stations in TeleForm®; print, bubble, hand score and scan 10 test sheets per station, per organization, per language and perform quality assurance (QA) checks of raw data in CSV files	
	QEII Production Coordinator	(Initials & sign-off date)
2.	Create answer keys, compare against exam sheets, case, scoring rules and adjust as necessary; create and QA post-encounter probe (PEP) answer keys and other docs for PEP marking	
	QEII Production Coordinator	(Initials & sign-off date)
3.	Enter answer key into SQL database and download answer key to SAS	
	Analyst (SQL)	(Initials & sign-off date)
	Analyst (SAS)	(Initials & sign-off date)
4.	Update the SAS code as necessary (new question type, etc.)	
	Analyst (SAS)	(Initials & sign-off date)
5.	Import CSV raw data into scoring application and run scoring on dummy data, run SAS scoring in parallel and verify matching results in SAS and scoring application	
	QEII Production Coordinator	(Initials & sign-off date)
	Analyst (SQL)	(Initials & sign-off date)
	Analyst (SAS)	(Initials & sign-off date)
	Senior Psychometrician (Overview of analyses)	(Initials & sign-off date)
6.	Test candidate bar code labels	
	QEII Production Coordinator	(Initials & sign-off date)
7.	Test examiner bar code labels	
	Administration Coordinator	(Initials & sign-off date)
8.	Prepare a file with demographic information about candidates including creating candidate groups (candidate info)	
	Analyst (SQL)	(Initials & sign-off date)
	Analyst (SAS)	(Initials & sign-off date)
	Senior Psychometrician (Overview of analyses)	(Initials & sign-off date)

Post-examination (Pre-CEC)

These tasks *MUST* be completed before the CEC meets.

9.	On-site scanning, receiving and uploading data files from sites and load files into TeleForm®	
	QEII Production Coordinator	(Initials & sign-off date)
10.	Verify	
	CSA team	(Initials & sign-off date)
11.	PEP marking, scan PEP sheets, upload files to TeleForm® and verify	
	Production Coordinator/Test Development Officer, MCCQE Part II	(Initials & sign-off date)
	Manager, MCCQE Part II	(Initials & sign-off date)
12.	Import CSVs into scoring application and correct until error reports are clean	
	QEII Production Coordinator	(Initials & sign-off date)
	Analyst (SQL)	(Initials & sign-off date)
13.	Run SAS scoring in parallel with scoring in SQL; verify matching results in SAS and scoring application	
	Analyst (SQL)	(Initials & sign-off date)
	Analyst (SAS)	(Initials & sign-off date)
14.	Run pull lists and QA against SAS	
	QEII Production Coordinator	(Initials & sign-off date)
	Production Coordinator, CSA	(Initials & sign-off date)
	Analyst (SAS)	(Initials & sign-off date)
15.	Pull sheets, hand score, enter into hand scoring application and update CSV files	
	QEII Production Coordinator	(Initials & sign-off date)
	Production Coordinator, CSA	(Initials & sign-off date)
16.	PEP QA	
	Production Coordinator/Test Development Officer, MCCQE Part II	(Initials & sign-off date)
	Manager, MCCQE Part II	(Initials & sign-off date)
17.	Run preliminary statistical analyses	
	Analyst (SAS)	(Initials & sign-off date)
18.	Review statistical information	
	Senior Research Psychometrician	(Initials & sign-off date)

19.	Present preliminary statistics to the Manager, MCCQE Part II and Associate Director, Evaluation Bureau	
	Senior Research Psychometrician	(Initials & sign-off date)
20.	If the statistical analyses indicate an unusual pattern, call a meeting with the Manager, MCCQE Part II, Associate Director, Evaluation Bureau, Senior Research Psychometrician and Analysts (SQL and SAS)	
	Senior Research Psychometrician	(Initials & sign-off date)
	<p>Note unusual pattern: * Add any supporting documentation to the folder. * Additional notes or comments can be added to the last page of this document.</p>	
21.	Hand score and data entry changes from PEP QA into hand score application; run error reports and correct errors to hand score data	
	Temp staff	(Initials & sign-off date)
	QEII Production Coordinator	(Initials & sign-off date)
22.	Enter data changes in CSV files from PEP QA, QA data entry and re-run scoring	
	National Site Coordinator/other MCCQE Part II staff	(Initials & sign-off date)
	QEII Production Coordinator	(Initials & sign-off date)
	Analyst (SQL)	(Initials & sign-off date)
23.	Run SAS scoring in parallel with scoring application; verify matching results in SAS and scoring application	
	Analyst (SAS)	(Initials & sign-off date)
	Analyst (SQL)	(Initials & sign-off date)
24.	Other QA reports and corrections (i.e., candidates with zero scores on components, examiner analyses, missing data, etc.)	
	QEII Production Coordinator	(Initials & sign-off date)
25.	Examiner feedback: data entry and editing	
	QEII Production Coordinator	(Initials & sign-off date)
26.	Incident reports: data entry, classification, and follow-up	
	Administration Coordinator (A)	(Initials & sign-off date)
	Administration Coordinator (B)	(Initials & sign-off date)
27.	Patient safety and lapses in professional behaviour	
	QEII Production Coordinator	(Initials & sign-off date)
28.	Special case investigations	
	Analyst (SQL)	(Initials & sign-off date)
	Analyst (SAS)	(Initials & sign-off date)
	Manager, Data Analyses (overview of Analysts)	(Initials & sign-off date)

	Senior Research Psychometrician (overview)	(Initials & sign-off date)
29.	Super borderline review	
	Production Coordinator and team, MCCQE Part II	(Initials & sign-off date)
	Manager, MCCQE Part II	(Initials & sign-off date)
30.	Ad-hoc investigations	
	QEII Production Coordinator	(Initials & sign-off date)
	Analyst (SQL)	(Initials & sign-off date)
31.	CEC teleconference	
	Manager, MCCQE Part II	(Initials & sign-off date)
32.	Prepare reports to CEC (special cases)	
	Manager, MCCQE Part II	(Initials & sign-off date)
33.	Update CSV files and re-score (if necessary)	
	QEII Production Coordinator	(Initials & sign-off date)
	Analyst (SQL)	(Initials & sign-off date)
34.	Re-run all quality assurance steps	
	Analyst (SQL)	(Initials & sign-off date)
	Analyst (SAS)	(Initials & sign-off date)
	Senior Research Psychometrician (final review)	(Initials & sign-off date)
35.	Prepare CEC report	
	Manager, MCCQE Part II	(Initials & sign-off date)
	Senior Research Psychometrician	(Initials & sign-off date)
36.	Dry run and approval of the CEC report	
	Associate Director, Evaluation Bureau	(Initials & sign-off date)
	Manager, MCCQE Part II	(Initials & sign-off date)
	Senior Research Psychometrician	(Initials & sign-off date)
37.	Test import into physiciansapply.ca (staging)	
	Analyst (SQL)	(Initials & sign-off date)
38.	QA SoR and SFR from staging environment	
	Manager, MCCQE Part II	(Initials & sign-off date)

CEC

39.	Present CEC report and obtain approval of results	
	Senior Research Psychometrician (review item)	(Initials & sign-off date)
	Manager, MCCQE Part II	(Initials & sign-off date)
40.	Note/document decisions by CEC	
	Manager, MCCQE Part II	(Initials & sign-off date)
41.	Update SQL and SAS with decisions by CEC	
	Analyst (SQL)	(Initials & sign-off date)
	Analyst (SAS)	(Initials & sign-off date)
	Senior Research Psychometrician (review item)	(Initials & sign-off date)
42.	Export physiciansapply.ca data from scoring application to Access database	
	Analyst (SQL)	((Initials & sign-off date)
43.	QA Access	
	Analyst (SAS)	(Initials & sign-off date)
44.	Import to physiciansapply.ca (production), validate and approve	
	Manager, MCCQE Part II	(Initials & sign-off date)
	Associate Director, Evaluation Bureau	(Initials & sign-off date)
	Director, Evaluation Bureau	(Initials & sign-off date)
45.	QA production SoR and SFR	
	Manager, MCCQE Part II	(Initials & sign-off date)
	Senior Research Psychometrician	(Initials & sign-off date)
46.	Updating the CEC report with the CEC decisions (last slide)	
	Manager, MCCQE Part II	(Initials & sign-off date)

APPENDIX B: PATIENT INTERACTION RATING SCALE ITEMS

1. Initiation of interview [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lack of introduction	Minimal acknowledgement of patient	Borderline unsatisfactory; acknowledges patient, introduces self	Borderline satisfactory; acknowledges patient, introduces self	Acknowledges patient; moderately at ease and attentive	Attentive to patient; introduces self; at ease, personable

2. Listening skills [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Interrupts inappropriately, ignores patient's answers	Impatient	Borderline unsatisfactory; somewhat attentive	Borderline satisfactory; somewhat attentive	Attentive to patient's answers	Consistently attentive to answers and concerns

3. Questioning skills [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Awkward, exclusive use of closed-ended or leading questions and jargon	Somewhat awkward; inappropriate terms; minimal use of open-ended questions	Borderline unsatisfactory; moderately at ease; appropriate language; uses different types of questions	Borderline satisfactory; moderately at ease; appropriate language; uses different types of questions	At ease; clear questions; appropriate use of open and closed-ended questions	Confident; skillful questioning

4. Organization of interview [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Scattered, shot-gun approach	Minimally organized	Borderline unsatisfactory; somewhat logical flow	Borderline satisfactory; logical flow	Logical flow with sense of purpose	Purposeful, integrated handling of encounter

5. Rapport with person [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Condescending, offensive, judgmental	Minimal courtesies only	Borderline unsatisfactory	Borderline satisfactory	Polite and interested	Warm, empathic

6. Information giving [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
No attempt or inappropriate attempt to give information (e.g., not truthful)	Awkward and / or incomplete attempts to give information	Borderline unsatisfactory; somewhat at ease, attempts to give information	Borderline satisfactory; somewhat at ease, attempts to give information	Gives information easily; somewhat attentive to patient's understanding	Confident and skillful at giving information; attentive to patient's understanding (e.g., truthful)

7. Professional behaviour with patient [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Offensive or aggressive; frank exhibition of unprofessional conduct	Negative attitude toward patient	Borderline unsatisfactory; does not truly instill confidence	Borderline satisfactory; manner inoffensive, but does not necessarily instill confidence	Attempts professional manner with some success	Overall demeanour of a professional; caring, listens, communicates effectively

8. Professional behaviour with colleague [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Offensive or aggressive; frank exhibition of unprofessional conduct	Negative attitude	Borderline unsatisfactory; not truly respectful	Borderline satisfactory; moderately respectful	Demonstrates professional manner with some success	Overall demeanour of a professional; listens, communicates effectively

9. Ethical conduct [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Markedly inappropriate or awkward handling of ethical issues	No consideration of ethical issues	Borderline unsatisfactory; minimal consideration of ethical issues	Borderline satisfactory; minimal and appropriate consideration of ethical issues	Responds satisfactorily to ethical issues	Considers ethical issues with care and effectiveness, and responds to them

10. Organization of physical examination [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exam not attempted or scattered; patient moved unnecessarily	Minimally organized	Borderline unsatisfactory; somewhat logical flow	Borderline satisfactory; logical flow	Logical flow with sense of purpose	Purposeful, integrated handling of examination

11. Attention given to patient's physical comfort [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exam not attempted, or inattentive to patient's comfort or dignity (e.g., no draping, causes pain unnecessarily)	Causes some unnecessary discomfort or embarrassment	Borderline unsatisfactory in attending to patient's comfort and needs	Borderline satisfactory in attending to patient's comfort and needs	Mostly attentive to patient's comfort and dignity	Consistently attentive to patient's comfort and dignity

12. Demonstration of technical skills [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exam not attempted or maneuvers cannot provide reliable / useful information	Maneuvers too rushed or clumsy; unlikely to provide reliable / useful information	Borderline unsatisfactory; some skill, but minimal likelihood of reliable / useful findings	Borderline satisfactory; some skill, some reliable / useful findings likely	Consistent skill; maneuvers likely to provide reliable / useful information	Consistent skill; maneuvers performed will elicit reliable / useful information

13. Relationship to the patient [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Introduction absent or inappropriate; no consent; awkward; uses jargon; no acknowledgment of patient	Minimal interaction and / or minimal acknowledgment of patient	Borderline unsatisfactory in approach to patient	Borderline satisfactory in approach to patient	Moderately clear and understandable; acknowledges patient; moderately at ease with patient	Clear, concise instructions; elicits consent to physical examination; at ease with patient

14. Overall organization of patient encounter [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
No logical flow; scattered, inattentive to patient's agenda	Counsels patient before taking history or doing physical	Minimal organization; scattered approach	Appropriate approach to patient	Skillful approach to patient	Skillful, professional approach to patient and effective use of time

15. Compliance optimization [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Candidate's approach may negatively affect compliance	Patient's compliance unlikely to be optimized	Borderline unsatisfactory; weak attempt to encourage patient's compliance	Borderline satisfactory; candidate's approach may positively affect patient's compliance	Candidate's approach encourages patient's compliance	Candidate's approach highly likely to optimize patient's compliance

16. Facilitation of informed decision making [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
No attempt or inappropriate attempt at information sharing (e.g., deception, slanting of facts, incorrect information)	Incomplete and / or biased information; overuses jargon; does not ensure understanding of issues	Attempts to share information; omits some critical facts; uses some jargon; attempts to ensure understanding	Gives some information on most important facts; may use jargon; attempts to ensure understanding	Gives clear information; supports patient decision making (e.g., alternatives, risks/benefits); appropriate language; ensures understanding	Organized; optimizes patient decision making; significant effort to make information relevant; clear language; attentive to patient understanding

17. Responds to question [ONE bubble only]

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ignores question or is rude, dismissive	Poor answer or explanation	Borderline unsatisfactory answer or explanation	Borderline satisfactory answer or explanation	Clear answer or explanation; is empathetic during interaction	Clear answer or explanation; empathetic; asks if understands or needs more information

APPENDIX C: EXAMPLE SOR AND SFR



Medical Council of Canada Qualifying Examination Part II Statement of Results

Candidate name: Vvvvvvvv, Vvvvvv Vvvvvvvv
Candidate code: 0000000000
Examination session: May 2018 **Your final result:** Pass
Pass score: 509 **Your total score:** 775

May 30, 2018

We are writing to inform you of your final result on the Medical Council of Canada Qualifying Examination Part II.

Your total score, representing your overall performance, is reported as a scaled score ranging from 50 to 950. Your final result is based on your total score relative to the pass score.

For more information, please visit the exam's Scoring web page on our website, mcc.ca.

Supplemental information on your examination performance is reported to you in a separate document within your physiciansapply.ca account.

mcc.ca
physiciansapply.ca
inscriptionmed.ca



SUPPLEMENTAL FEEDBACK REPORT

1021 Thomas Spratt Place
1021, place Thomas Spratt
Ottawa, ON
Canada K1G 5L5
613-521-6012

Candidate name: Vvvvvvvv, Vvvvvv Vvvvvvvv

MCC candidate code: 0000000000

Examination: MCCQE Part II

Your total score: 775

Examination session: May 2018

The purpose of this report is to provide you with supplemental information on your relative strengths and weaknesses, based on your performance across the different domains that were assessed by the test form of the Medical Council of Canada Qualifying Examination (MCCQE) Part II that was administered to you.

Figure 1 displays your performance measured across stations in the following four domains: C²LEO (Considerations for Cultural-Communication, Legal, Ethical, and Organizational aspects of the Practice of Medicine), Data Acquisition, Patient / Physician Interaction, and Problem Solving and Decision Making. Each domain is sampled a number of times, with some being measured across all stations and others across a smaller number of stations.

- The C²LEO subscore is calculated using checklist items, oral questions, rating scales, and/or written questions across the stations that measure the C²LEO domain.
- The Data Acquisition subscore is calculated using items within stations that pertain to history taking and physical examination.
- The Patient / Physician Interaction subscore is calculated using rating scale items within most of the stations.
- The Problem Solving and Decision Making subscore is calculated using oral question items, relevant items within the management stations, as well as items from the written stations.

To help you better understand your performance, your subscore for each domain is shown along with the mean score of candidates from this examination session who were first-time takers of the MCCQE Part II and who passed. The standard error of measurement (SEM) associated with each of your subscores represents the expected variation in your subscore if you were to take this examination again with a different set of stations covering the same or similar domains. Small differences in subscores or overlap between SEMs are indicative that performance in those domains was relatively similar. Likewise, overlap between the SEM for a domain subscore and the mean score of first-time takers who passed, within a given domain, signifies that performance is similar to the mean.

Report date: 2018-05-30

1/2

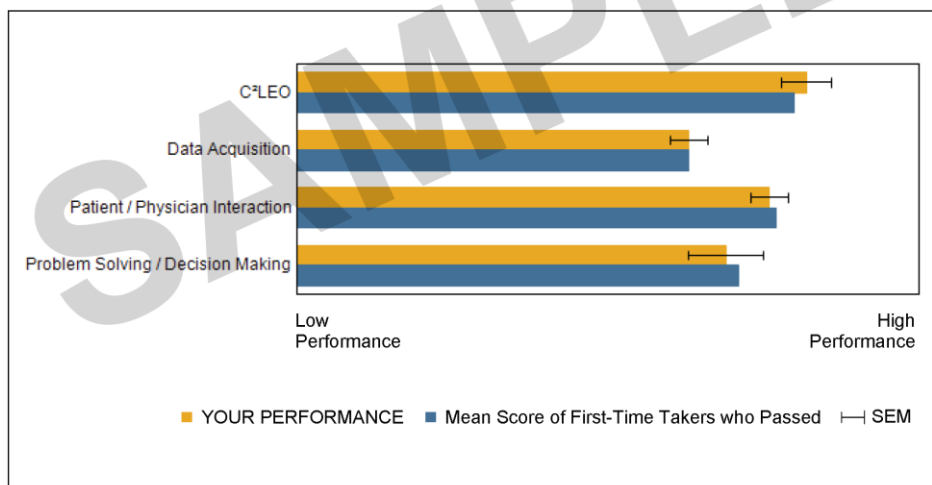
mcc.ca
physiciansapply.ca
inscriptionmed.ca

SUPPLEMENTAL FEEDBACK REPORT

It is important to note that the subscores are based on significantly less data than the total score and that these do not have the same level of precision as the total score. If you have failed the examination and wish to retake it, preparation for all domains is important; otherwise you could improve some subscores and inadvertently lower others. Please also note that your subscores cannot be directly compared to your total score in that they are calculated differently and as such the pass score cannot be translated to Figure 1.

For more information, please visit the [MCCQE Part II web page](#).

Figure 1. MCCQE Part II Score Profile



Report date: 2018-05-30

2/2

Vvvvvvvv, Vvvvvv Vvvvvvv / 0000000000

mcc.ca
physiciansapply.ca
inscriptionmed.ca