MEDICAL COUNCIL
OF CANADA

LE CONSEIL MÉDICAL
DU CANADA

Technical Report
on the Standard Setting
Exercise for the
Medical Council of Canada
Qualifying Examination
Part II

Psychometrics and Assessment Services
July 2015

PURPOSE OF THE MEETING

Background

A passing score, irrespective of any particular examination, should be revalidated every three to five years to ensure that the standard is still appropriate. A rigorous and valid process to establish the cut score should in particular be adhered to for licensing examinations (Cizek, 2012). This report outlines the processes, procedures and results of a standard setting exercise carried out for the Medical Council of Canada's Qualifying Examination Part II (MCCQE Part II). On February 23-25, 2015, 20 physicians from across Canada met at the Medical Council of Canada (MCC) office in Ottawa to participate in a three-day standard setting exercise that led to the recommendation of a passing score for the MCCQE Part II examination. The MCCQE Part II is a national, standardized examination that assesses the core fundamental knowledge, skills and attitudes expected of all physicians, regardless of specialty, essential for medical licensure in Canada prior to entering independent practice. The MCCQE Part II is composed of a series of Objective Structured Clinical Examination (OSCE) stations that may include, but are not limited to, problems in medicine, pediatrics, obstetrics, gynecology, psychiatry and surgery.

Our standard setting exercise resulted in a recommended passing score for consideration by the Central Examination Committee (CEC). Panelists were informed that their role was one of recommending a passing score rather than setting a passing score. Final approval of the recommended passing score is the responsibility of the CEC; a body which is responsible for the oversight of the MCCQE Part II, including the approval and maintenance of exam content and approval of exam results.

PROCEDURES

The present standard setting exercise was preceded by a review of potential methods and related issues to consider for setting passing scores on exams such as the MCCQE Part II that is composed of OSCE stations. In evaluating standard setting methods that are appropriate for the MCCQE Part II, taking into account the multidimensional nature and complexity of OSCE stations, two methods were considered for this standard setting exercise: the contrasting groups and borderline group methods. A modified version of the borderline group method had been used to determine the cut score on the MCCQE Part II exam from its inception until fall 2012. This method used physician examiners' global rating judgments provided while scoring the examination to set the standard for each examination. The National Assessment Collaboration Exam used the borderline group method in a panel based standard setting exercise in March 2013. The CEC had been given information on these activities and endorsed our recommendation that the borderline group method be used for the spring 2015 standard setting exercise for the sake of consistency across OSCEs.

Planning of the standard setting exercise, as well as the review of materials and documents, was conducted by two MCC psychometricians. Other MCC staff supported the preparation and delivery of the standard setting exercise. In the remainder of this section, we present a description of how the panelists were selected, the information provided to the panelists prior to the three day meeting, the method used to set the passing score, and a description of the events that took place during the three-day meeting.

Selecting Panelists

Many features of a standard setting exercise can influence the validity of the recommended passing score as well as its associated process. One of these features is the selection of well-qualified panelists. In October 2014, the MCC sent an email to physician test committee members and physician examiners soliciting participation in our standard setting exercise, which resulted in more than 50 physicians being nominated. Each nominee completed a demographic information form. The original invitation email and demographic information form are shown in Appendix A.

On the basis of the demographic information collected, MCC staff selected 20 participants with the intent to create two matched panels of ten panelists each, denoted as subpanel 1 and subpanel 2 in Table 1. While a multitude of background information was collected, we focused the assembly of the two panels using the variables listed in Table 1. Every effort was made to

match both panels as closely as possible on the following key variables: gender, geographic region, ethnic background, medical specialty, and number of years in practice.

Table 1: Demographic Information by Standard Setting Subpanel

Variable of Interest	Group	Subpanel 1	Subpanel 2	Total
Gender	Male	4 (40%)	5 (50%)	9 (45%)
	Female	6 (60%)	5 (50%)	11 (55%)
Geographic region	West	2 (20%)	2 (20%)	4 (20%)
	Prairies	1 (10%)	1 (10%)	2 (10%)
	Ontario	4 (40%)	4 (40%)	8 (40%)
	Quebec	2 (20%)	2 (20%)	4 (20%)
	Atlantic Prov.	1 (10%)	1 (10%)	2 (10%)
Ethnic background	Caucasian	6 (60%)	7 (70%)	13 (65%)
	Asian	4 (40%)	2 (20%)	6 (30%)
	Black	0 (0%)	1 (10%)	1 (5%)
Specialty	Primary care	6 (60%)	5 (50%)	11 (55%)
	Other care	4 (40%)	5 (50%)	9 (45%)
Number of years in practice post-residency	1 to 10	5 (50%)	6 (60%)	11 (55%)
	11 to 30	4 (40%)	3 (30%)	7 (35%)
	30+ years	1 (10%)	1 (10%)	2 (10%)

Assigning Panelists to Two Panels

Panelists were assigned to one of the two panels on the first morning of the standard setting exercise. The primary purpose of having two panels was to allow MCC staff to assess the generalizability of the passing score across both matched groups. Furthermore, smaller panels foster more discussion amongst members. If a panel is too big, it becomes more difficult for individual panelists to share their views due to competing availability of time and other group dynamic factors. Splitting panelists into groups tempers this concern.

Additionally, having two panels allows us to assess the generalizability of the passing score across groups (i.e., can we replicate the passing scores across two matched panels?). Demonstrating this comparability across two independent panels lends considerable credibility to the ensuing passing score. When panel passing scores are highly related, they are usually averaged to produce a passing score after each round. Conversely, in instances where they diverge, a discussion among panelists can follow to clarify any reasons that might account for this discrepancy (e.g., the two panels simply had very different ideas, or one or two panelists

station for potentially all panelists. To ensure data integrity, we decided to recollect Round 2 ratings for all 12 stations from all panelists (c.f. Round 2b next).

Round (2b)

The meeting proceeded with the collection of each panelist's independent judgments for each of the stations in Round 2b. Each panelist was provided their ratings for Round 1 and the ratings they had completed for nine to ten stations from Round 2a on paper. Given some of the challenges with the electronic data capture tool, we asked panelists to also write down their Round 2b ratings on their printed copies. Following Round 2b, the standard setting exercise proceeded as planned with a gathering of Hofstee data and a presentation of the Round 2b MCCQE Part II passing score.

Quality Assurance

As a quality assurance (QA) measure, post standard setting, we compared the results of the online ratings to the paper ratings that each of the panelists provided due to our concern of panelists' fatigue and the resulting potential of data entry errors. The paper ratings were entered twice into an excel sheet independently and compared to the online ratings provided in Round 2b. One hundred and eighty-nine (1.6%) ratings (out of 12,000 total ratings) were not consistent between Round 2b and QA entries. Most of the inconsistencies in ratings were due to two panelists; 24 from one panelist, and 68 from a second panelist. The second panelist had the ratings for one station captured incorrectly, where the ratings for station C07 and C09 were identical. Based on our evaluation, we deemed the QA paper entries to be more accurate. Thus, we recommended that the passing score resulting from our QA check be adopted by the CEC. In this report, Round 1 and Round 2b results are presented along with the recommended passing score from the QA process.

Incorporating political and other considerations: The Hofstee Method

Prior to concluding each round, we asked panelists to answer four specific questions which define the Hofstee method. The latter is generally viewed as a procedure which allows judges to gauge the appropriateness of standards in light of a reality check which includes both criterion-referenced (acceptable cut score) and norm-referenced (acceptable failure rate) considerations. A description of the method was presented to the group as a whole followed by the entry of their judgments on paper (see Appendix D). Specifically, panelists were asked to specify the lowest and highest passing scores that they believed were reasonable for the MCCQE Part II exam. Additionally, panelists were asked to provide the lowest and highest failure rates that they felt were tolerable. Panelists provided acceptable low and high passing score values on the actual

conditions that the test user would be equally willing to accept (Shavelson & Webb, 1991). G-theory provides a summary coefficient reflecting the level of dependability (D-coefficient) and a generalizability coefficient (G-coefficient) that is analogous to classical test theory's reliability coefficient. Multiple sources (commonly called facets) of error in a measurement, can be estimated separately in a single analysis e.g., persons or candidates, items (or in the case of OSCEs, stations), raters or subpanel. The purpose of our analyses was to determine how much variance was attributable to sources that are undesirable, such as raters, subpanels, and stations and how much variance was due to actual differences in candidate abilities (true score variance, which is desirable in an effort to separate passing from failing candidates).

We conducted a G-study with three facets (station, rater and subpanel) in a *person x station x (rater: subpanel)* design. In other words, the same 50 candidates were rated on the same stations by panelists who were nested (assigned) to a specific sub-panel. The ratings obtained from the QA process were used for these analyses. Table 3 shows the variance components for the candidates' ratings as well as each source of possible measurement error. The largest facet, not surprisingly, was the person x station interaction which accounted for 47.9% of the total variance. This indicates that the performance of candidates (on the 1-3 scale) varied by station. This is commonly referred to as case specificity (Norman, Bordage, Page & Keane, 2006), which implies that success on any case or station is specific to that case and does not generalize very well to other stations. This is a common occurrence in OSCEs due to the smaller number of stations that can be realistically administered in an exam form (as compared to MCQs, for example). The second largest effect was noted for the person facet (13.1% of total variance), which indicates that candidates did differ in their overall ability. This is akin to true score variance and suggests that the MCCQE II was able to separate out candidates, in terms of their ability level. The third largest effect was reported for the station facet which accounted for 8.5% of the total score variance. This suggests that stations differed in their overall difficulty level.

Because the raters (or panelists) were nested within each subpanel, the rater effect cannot be interpreted without the associated nested component of panel. The rater- related effects were the next group of facet effects that were examined: rater: panel accounting for 1.1% of total variance; station x (rater: panel) explaining 4.1% of total variance and; person x (rater: panel), accounting for 0.2% of total rating variance. These results indicate that about 1.1% of the total rating variance was due to the rater nested within the panel. In other words, the cut score was nearly identical across raters.

The panel related effects were the next group of effects that were examined: panel accounted for 0.1% of the total rating variance; whereas the person x panel and station x panel effects

accounted for essentially no rating variance. These results indicate that there was a negligible amount of variance due to the two subpanels. These findings indicate that: (a) the cut score was nearly identical, irrespective of subpanel. The G- coefficient and D-coefficient for this model was 0.76 and 0.73 respectively, which indicates that the ratings provided for this standard setting exercise would generalize quite well if a different set of candidates, raters or subpanels were to be used. These results would generalize less well if a different set of stations were to be used since most of the variance is associated with person x station, which indicates that the cut score established for this exam is dependent on the set of stations used to set the standard and would necessitate that test score linking be implemented to ensure comparability of this standard across test forms (Kolen & Brennan, 2004). Relating to this point, please note that we have implemented test score linking, as of the spring 2015 administration, for the MCCQE Part II examination.

Table 3: Results of Generalizability Theory Variance Component Estimates

Facet	df	SS	EMS	EVC	% Variance
person	49	1318.0	26.9	0.1	13.1
station	11	696.6	63.3	0.1	8.5
panel	1	9.1	9.1	0.0	0.1
person x station	539	3449.1	6.4	0.3	47.9
person x panel	49	7.6	0.2	0.0	0.0
station x panel	11	9.6	0.9	0.0	0.0
person x station x panel	539	74.3	0.1	0.0	0.0
rater: panel	18	101.0	5.6	0.0	1.1
person x (rater: panel)	882	157.5	0.2	0.0	0.2
station x (rater: panel)	198	306.8	1.6	0.0	4.2

df = degrees of freedom SS = sums of squares
EMS = Expected Mean Squares
EVC = Estimated Variance Components
% Variance = Percentage of Total Variance

Impact Data – Pass Rates

The results of the impact on candidate groups are shown in Table 4. The pass rate for Round 1, Round 2b and the QA ratings are shown for Canadian Medical Graduates, registered as Canadian Postgraduate first time test takers (CMG-CPG 1st), first time test takers, and all candidates (or total) for the MCCQE Part II Fall 2014 test form. The overall pass rate is lower for Round 2b and the QA ratings as compared to Round 1 as the cut score increased between Round 1 and Round 2b and between Round 1 and the QA ratings.

Table 4: Pass Rates by Round and Candidate Cohort for Fall 2014 Exam

Candidate Cohort	Round 1 61.86	Round 2b 63.22	QA ratings 63.06
CMG-CPG first time test taker	89.92	85.69	86.12
First time test taker	81.81	76.35	76.73
Total	75.60	69.70	70.37

Hofstee Results

The Hofstee results were computed for each panel as a function of round (Round 1 and Round 2b; see Table 5). Round 1 and QA version ratings were similar across subpanels, with the exception of the maximum failure rate where subpanel 2 expectations led to a higher maximum failure rate. There were slight differences between Round 1 and Round 2b results within each subpanel. The main differences were noted in the maximum and minimum failure rates. All of the ranges provided by the panelists fall within the borderline group cut scores shown in Table 2. This indicates that the panelists' "gut" estimates were in line with the results based on the borderline group method.

Table 5: Summary of Hofstee Results by Round and Panel

Round	Statistic	Subpanel 1	Subpanel 2	Across Panels
Round 1	Percent Min	56.50	53.20	54.85
	Percent Max	71.00	70.90	70.95
	Failure Min	6.70	7.20	6.95
	Failure Max	35.50	48.50	42.00
Round 2b	Percent Min	56.90	52.30	54.60
	Percent Max	71.80	70.70	71.25
	Failure Min	6.05	4.50	5.28
	Failure Max	29.00	43.00	36.00

Summary of Evaluation Survey Findings

The evaluation survey was divided into sections that largely reflect major activities that occurred over the three-day meeting. See Appendix E for a full summary of the survey across all panelists and by subpanel with each survey question and results presented[1]. Overall findings, of the

---

[1] One panelist left the meeting after the second round of impact data was presented and did not complete the survey.

The Hofstee results provided a "gut" check that the passing score established by subpanel and across panels was within acceptable ranges, based on an overall holistic impression. The Hofstee results for both Round 1 and Round 2b provided boundaries that were in line with the panelists ratings for the borderline group method, as well as resulting pass rates that would ensue based on the Fall 2014 MCCQE Part II form; note that the latter was not that disparate from historical pass rates.

Finally, the results of the survey conducted at the end of the three-day standard setting exercise were quite positive, indicating that the experience from the panelists' point of view was excellent and that we achieved our intended goals of preparing the panelists appropriately. Ultimately, and most importantly, panelists were very confident in the recommended passing score. These results are similar to those found with other standard setting exercises, including our MCCQE Part I exam, and NAC exam. Ultimately, the survey results provide additional validation evidence in support of the recommended passing score being proposed to the CEC.

In summary, the similarity of the cut scores by panel, generalizability results, Hofstee results, impact data being similar to past administrations, and survey results all provide evidence that this standard setting exercise was validated appropriately. The panel- based standard setting exercise was a thorough and rigorous process in establishing a passing score and met best practice standards and procedures.

The CEC was presented the information in this report and impact information for applying this new cut score to the spring 2015 candidate results. Using the spring 2015 results of all MCCQE Part II candidates, the new scale was established to have a mean of 500 and a standard deviation of 100. On this new scale, the pass score that was recommended from the standard setting panel and approved by the CEC is 509. This pass score will remain in place for subsequent MCCQE

REFERENCES

Boulet J., De Champlain, A. F. & McKinley, D. (2003). Setting defensible performance standards on OSCEs and standardized patient examinations. Medical Teacher, 25, 245-9.

Cizek, G. J. (2012). An introduction to contemporary standard setting: Concepts, characteristics, and contexts. In G. J. Cizek (Ed.), Setting performance standards: Foundations, Methods, and Innovations (pp. 3-14). New York, NY: Routledge.

De Champlain, A. F. (2013). Standard setting methods in medical education. In T. Swanwick (Ed.). Understanding Medical Education: Evidence, Theory and Practice. (pp. 305-316). Chichester, West Sussex: John Wiley & Sons, Ltd.

De Champlain, A. F. (2004). Ensuring that the competent are truly competent: An overview of common methods and procedures used to set standards on high-stakes examinations. Journal of Veterinary Medical Education, 31, 61-5.

Kolen, M. J., & Brennan, R. L. (2004). Test equating, scaling, and linking: Methods and practices (2nd ed.). New York, NY: Springer Science + Business Media, LLC.

Norman, G., Bordage, G., Page, G., & Keane, D. (2006). How specific is case specificity? Medical Education, 40 (7), 618-23.

Shavelson, R. J., Webb, N. M. (1991). Generalizability Theory: A Primer. Sage Thousand Oaks, CA: Publications Inc.

APPENDIX A:
Invitation letter and demographic sheet

Dear Prospective Panelist:

In an effort to set the performance standard for Medical Council of Canada's Qualifying Examination Part II (MCCQE Part II), the governing bodies of the Medical Council of Canada (MCC) have decided to launch a standard setting exercise. To begin this process, the Research and Development directorate at the MCC is soliciting participation for a panel to recommend passing scores. It is expected that the final passing score will be used for the examination starting with the spring 2015 administration.

We hope that you will consider volunteering to participate on our panel, as your clinical expertise and past experience are vital to the success of this standard setting exercise. We are sending out this notice to solicit volunteers from which we will assemble the panel to ensure that the diversity of medical experts and clinical practice contexts across Canada are well represented.

Selected panelists will carry out the review task on February 23-25, 2015 at the MCC offices in Ottawa. Panelists will be trained to evaluation examination materials and will be guided through a set of procedures to set the passing score. An honorarium of $500 per day (full 3-day meeting) plus reasonable travel and accommodation expenses will be provided.

Should you be interested in participating, we ask that you fill out the attached Demographic Information Sheet, return it to the MCC tentatively reserve the dates of February 23-25, 2015, in your calendar. Responses are requested by October 15, 2014 and your participation will be confirmed by November 19, 2014.

Thank you very much for your interest and support in achieving the highest standards in the assessment of Medical Graduates in Canada.

Sincerely,

Acting Director
Research & Development
Medical Council of Canada

Demographic Information Sheet

The information requested below is being collected to help the MCC obtain a pan- Canadian representative panel to recommend a passing score on the MCCQE Part II Examination. This information will only be used to select the panel members so that we can represent the diversity of physicians across the country. The information will not be linked in any way to the collection of data for setting the passing score. A reminder that the meeting will take place on February 23, 24, and 25, 2015 therefore we are asking panelists to be available on all three days.

Please provide your name and contact information, and check a box next to each of the questions. The form can be sent electronically to research@mcc.ca by October 15, 2014.

Name: _____

Email: _____ Phone number: _____

Mailing address: _____

_____

_____

1. Number of years in practice post residency:

1-5 years	☐
6-10 years	☐
11-20 years	☐
21-30 years	☐
More than 30 years	☐

2. Number of years' experience supervising residents:

1-5 years	☐
6-10 years	☐
11-20 years	☐
21-30 years	☐
More than 30 years	☐

3. **Do you have experience supervising Canadian Medical Graduates:**

      Yes   ☐

      No   ☐

4. **Have you ever been a member of a Medical Council test committee?**

      Yes   ☐

      No   ☐

      4b. If so, which test committee? _____

5. **Have you ever been an examiner for an Objective Structured Clinical Examination:**

      Yes   ☐

      No   ☐

6. **Country of medical training (post graduate training):**

      Canada   ☐

      Other   ☐

7. **Region of the country in which you live:**

      Alberta   ☐

      British Columbia   ☐

      Manitoba   ☐

      Maritimes   ☐

      Ontario   ☐

      Quebec   ☐

      Saskatchewan   ☐

      Territories   ☐

8. **First Language:**

      English   ☐

      French   ☐

      Other (_____) ☐

9. **Gender:**

Male	☐
Female	☐

10. **Ethnicity:**

Asian	☐
Black	☐
Caucasian	☐
First Nations	☐
Hispanic	☐

11. **Medical Specialty:**

Pediatrics	☐
Internal Medicine	☐
Psychiatry	☐
Obstetrics and Gynecology	☐
Surgery	☐
Family Medicine	☐
Other	☐

12. **Type of community in which you work:**

Urban	☐
Rural	☐

13. **Type of care setting:**

Hospital-based	☐
Community-based	☐

AGENDA – Tuesday, February 24, 2015		
08:00 a.m.	Continental breakfast at MCC (meeting begins at the same time)	
08:00 a.m.	Split into subpanels (round 1) - continued	Staff/Panelists
09:30 a.m.	Training for Couplet Station	TDO/Staff
10:15 a.m.	**BREAK**	
10:30 a.m.	Split into subpanels (round 1) - continued	Staff/Panelists
12:00 p.m.	**LUNCH**	
13:00 p.m.	Split into subpanels (round 1) - continued	Staff/Panelists
14:30 p.m.	Hofstee judgements	Psychometricians
14:40 p.m.	**BREAK**	
15:10 p.m.	Impact data	Psychometricians
15:55 p.m.	Split into subpanels (final round)	Staff/Panelists
16:55 p.m.	Wrap-up/overview of day 3	Psychometricians
17:00 p.m.	**End of day 2**	

AGENDA – Tuesday, February 24, 2015		
08:00 a.m.	Continental breakfast at MCC (meeting begins at the same time)	
08:00 a.m.	Split into subpanels (final round) - continued	Staff/Panelists
10:00 a.m.	**BREAK**	
10:15 a.m.	Split into subpanels (final round) - continued	Staff/Panelists
12:15 p.m.	**LUNCH**	
13:15 p.m.	Expense claims, taxis, etc.	Executive Assistant
13:30 p.m.	Split into subpanels (final round) - continued	Staff/Panelists
14:30 p.m.	Hofstee judgements	Psychometricians
14:40 p.m.	**BREAK**	
15:20 p.m.	Impact data	Psychometricians
15:30 p.m.	Post standard setting exercise survey	Psychometricians
16:30 p.m.	**End of day 3**	

APPENDIX C:
Description of the just qualified or borderline candidate

MCCQE Part II Standard Setting Exercise February 23-25, 2015		
Poor/unacceptable MCCQE Part II candidate	***/just qualified* MCCQE Part II candidate**	***Acceptable/good* MCCQE Part II candidate**
The candidate is not qualified for independent practice; their performance is poor/unacceptable. The MCCQE Part II candidate will usually demonstrate incomplete or disorganized data gathering from the history, physical examination and/or laboratory data. Even when sufficient information is gathered, the approach will be disorganized, the physical examination technique will be poor. AND/OR Interpretation of information obtained will not allow for a coherent differential diagnosis or management plan to be developed. AND/OR Interpersonal skills will be poorly demonstrated with little ability to engage the patient, will not be patient-centered and will not be sensitive to the patient's needs and understanding. The candidate will appear to lack confidence or be over-confident during the interaction with the patient.	The candidate is qualified for independent practice, but their performance is minimally acceptable. The MCCQE Part II borderline pass candidate will demonstrate an ability to gather some of the essential information (including laboratory data) about the patient and perform a physical examination that may lack some technical skill or be disorganized. Diagnostic information obtained will be minimally sufficient to allow the candidate to develop the expected differential diagnosis or management plan. AND Interpretation of information will be minimally organized for presentation and either some information will be overlooked or partly incorrect. AND Interpersonal skills will be minimally demonstrated. While information will be obtained from or provided to the patient/others, the approach will be minimally patient-centered. The candidate will be inconsistently responsive to verbal and non-verbal cues from the patient in relation to the patient's understanding of information being provided. The deficiencies will be such that the patient is not put at risk and the basic needs of the patient are met.	The candidate is qualified for independent practice. The acceptable MCCQE Part II candidate will demonstrate an ability to gather data of sufficient breadth and depth. This information provides the candidate the ability to develop a clear definition of a patient's problem through history gathering, a logical physical examination, and appropriate interpretation of laboratory data. AND Interpretation of information will be organized in a logical manner. The candidate can prioritize the information and reliably make an appropriate differential diagnosis. Based on the diagnosis, the acceptable candidate will consistently provide appropriate management. AND Interpersonal skills of the candidate will put the patient at ease, consistently showing respect, will demonstrate a patient-centered approach in gathering and providing information, and will verify the patient's understanding of any information provided.

Appendix D: Hofstee Paper Form

Panelist: _____

Subpanel: _____

Round: Initial

1. What is the **highest** percent passing score that would be acceptable, even if every candidate attains that score?

2. What is the **lowest** percent passing score that would be acceptable, even if no candidate attains that score?

3. What is the **maximum** acceptable failure rate?

4. What is the **minimum** acceptable failure rate?

Round: Final

1. What is the **highest** percent passing score that would be acceptable, even if every candidate attains that score?

2. What is the **lowest** percent passing score that would be acceptable, even if no candidate attains that score?

3. What is the **maximum** acceptable failure rate?

4. What is the **minimum** acceptable failure rate?

Appendix E:
Summary of responses to post-meeting survey

All Panelists

1. Which panel did you participate in? (Select ONE)

Response	Chart	Percentage	Count
Panel 1 (University room)		47.4%	9
Panel 2 (Barr/Bérard room)		52.6%	10
	Total responses		**19**

2. How clear were you about the description of the "Just Qualified" (or "Borderline Passing") candidate on the MCCQE Part II as you began the task of setting a passing score following the training on Monday morning? (Select ONE)

Response	Chart	Percentage	Count
Very clear		31.6%	6
Clear		42.1%	8
Somewhat clear		26.3%	5
Not clear		0.0%	0
	Total responses		**19**

3. Did you feel the discussion of the "Just Qualified" (or "Borderline Passing") candidate on the MCCQE Part II was helpful during the training on Monday morning? (Select ONE)

Response	Chart	Percentage	Count
Yes, very helpful		52.6%	10
Yes, helpful		42.1%	8
Yes, somewhat helpful		5.3%	1
Not helpful at all		0.0%	0
	Total responses		**19**

4. How would you judge the length of time spent (approximately 45 minutes) introducing, discussing, and editing the definition of the "Just Qualified" or "Borderline Passing" candidate? (Select ONE)

Response	Chart	Percentage	Count
About Right		89.5%	17
Too Little Time		5.3%	1
Too Much Time		5.3%	1
	Total Responses		**19**

5. What is your impression of the length of time for training you received for setting a passing score on the MCCQE Part II? (Select ONE)

Response	Chart	Percentage	Count
Appropriate		73.7%	14
Somewhat appropriate		26.3%	5
Not appropriate		0.0%	0
	Total Responses		**19**

6. How clear did you find the information that was provided regarding the scoring procedures for the MCCQE Part II? (Select ONE)

Response	Chart	Percentage	Count
Excellent		26.3%	5
Very good		47.4%	9
Good		21.1%	4
Fair		5.3%	1
Poor		0.0%	0
	Total Responses		**19**

7. What is your overall evaluation of the training that was provided for setting a passing score on the MCCQE Part II? (Select ONE)

Response	Chart	Percentage	Count
Excellent		26.3%	5
Very good		52.6%	10
Good		21.1%	4
Fair		0.0%	0
Poor		0.0%	0
	Total Responses		**19**

8. What factors influenced the ratings you made of "Just Qualified" (or "Borderline Passing") candidate responses on the MCCQE Part II? (Select ALL choices that apply)

Response	Chart	Percentage	Count
The description of the "Just Qualified" or "Borderline Passing" candidate		94.7%	18
My perception of the difficulty of the stations or station components		63.2%	12
The scoring of the individual stations or station components		57.9%	11
The station statistics (e.g., candidate station scores)		21.1%	4
The statistical impact data provided before round 2		31.6%	6
Panelists discussion		36.8%	7
My experience in the field		73.7%	14
Knowledge and skills measured by the stations		68.4%	13
Other (please specify)		0.0%	0
Total Responses			**19**

9. What factors influenced the ratings you made of "Just Qualified" (or "Borderline Passing") candidate responses on the MCCQE Part II? (Select ALL choices that apply)

Response	Chart	Percentage	Count
About Right		84.2%	16
Too Little Time		0.0%	0
Too Much Time		15.8%	3
Total Responses			**19**

10. Overall, how did you feel about participating in group discussions conducted during the ratings process for each station? (Select ONE)

Response	Chart	Percentage	Count
Very Comfortable		84.2%	16
Somewhat Comfortable		10.5%	2
Unsure		5.3%	1
Somewhat Uncomfortable		0.0%	0
Very Uncomfortable		0.0%	0
Total Responses			**19**

11. What level of confidence do you have that the impact data and final discussion on the final afternoon helped the panel arrive at a defensible passing score? (Select ONE)

Response	Chart	Percentage	Count
Very Confident		47.4%	9
Confident		42.1%	8
Somewhat Confident		10.5%	2
Not At All Confident		0.0%	0
	Total Responses		**19**

12. What level of confidence do you have in the final recommended passing score? (Select ONE)

Response	Chart	Percentage	Count
Very Confident		36.8%	7
Confident		47.4%	9
Somewhat Confident		15.8%	3
Not at All Confident		0.0%	0
	Total Responses		**19**

13. How could the method used for setting a passing score on the MCCQE Part II have been improved?

#	Response
1.	Good method
2.	This was excellent
3.	Not sure
4.	Less documents to be given on statistics, more on the subtleties of a Just passed score (there was only one explanation sheet)
5.	I know that 3 days is a long time! but having a bit more discussion on the key points of each station would be helpful
6.	No suggestions
7.	Problems were mostly unrelated to planning (e.g., room temperature, IT glitches)
8.	Consider having us do some actual scoring during the training sessions
9.	It would be informative to better understand how standardized scores are produced by the examiners on the day of the exam (i.e., how to translate an encounter into a standard score)
10.	Maybe more clarification on the criteria needed to grade a 3 - good

14. Please provide any additional comments or suggestions about the setting of a passing score on the MCCQE Part II.

#	Response
1.	I hope that with the multiple inputting of scores and participant fatigue that you got usable data. I do think that doing a couple of scores ahead of doing the marking would help before we dive into doing the 50 participants.
2.	Really enjoyed the work, happy to be involved
3.	I have no comments regarding this exercise. Having been an examiner and now having seen how the exam is scored, I believe that examiners should be better trained to mark the exam appropriately. I had never realized the importance of the second page (behaviour; attitude, etc)
4.	IMPORTANT COMPONENTS SHOULD BE DONE TO PASS A STATION
5.	Thank you

Subpanel 1

1. How clear were you about the description of the "Just Qualified" (or "Borderline Passing") candidate on the MCCQE Part II as you began the task of setting a passing score following the training on Monday morning? (Select ONE)

Response	Chart	Percentage	Count
Very Clear		33.3%	3
Clear		55.6%	5
Somewhat Clear		11.1%	1
Not Clear		0.0%	0
Total Responses			**9**

2. Did you feel the discussion of the "Just Qualified" (or "Borderline Passing") candidate on the MCCQE Part II was helpful during the training on Monday morning? (Select ONE)

Response	Chart	Percentage	Count
Yes, Very Helpful		55.6%	5
Yes, Helpful		44.4%	4
Yes, Somewhat Helpful		0.0%	0
Not Helpful at All		0.0%	0
Total Responses			**9**

3. **How would you judge the length of time spent (approximately 45 minutes) introducing, discussing, and editing the definition of the "Just Qualified" or "Borderline Passing" candidate? (Select ONE)**

Response	Chart	Percentage	Count
About Right		88.9%	8
Too Little Time		11.1%	1
Too Much Time		0.0%	0
	Total Responses		**9**

4. **What is your impression of the length of time for training you received for setting a passing score on the MCCQE Part II? (Select ONE)**

Response	Chart	Percentage	Count
Appropriate		100.0%	9
Somewhat appropriate		0.0%	0
Not appropriate		0.0%	0
	Total Responses		**9**

5. **How clear did you find the information that was provided regarding the scoring procedures for the MCCQE Part II? (Select ONE)**

Response	Chart	Percentage	Count
Excellent		33.3%	3
Very good		66.7%	6
Good		0.0%	0
Fair		0.0%	0
Poor		0.0%	0
	Total Responses		**9**

6. **What is your overall evaluation of the training that was provided for setting a passing score on the MCCQE Part II? (Select ONE)**

Response	Chart	Percentage	Count
Excellent		33.3%	3
Very good		55.6%	5
Good		11.1%	1
Fair		0.0%	0
Poor		0.0%	0
	Total Responses		**9**

7. **What factors influenced the ratings you made of "Just Qualified" (or "Borderline Passing") candidate responses on the MCCQE Part II? (Select ALL choices that apply)**

Response	Chart	Percentage	Count
The description of the "Just Qualified" or "Borderline Passing" candidate		88.9%	8
My perception of the difficulty of the stations or station components		77.8%	7
The scoring of the individual stations or station components		66.7%	6
The station statistics (e.g. candidate station scores)		22.2%	2
The statistical impact data provided before round 2		44.4%	4
Panelists discussion		55.6%	5
My experience in the field		77.8%	7
Knowledge and skills measured by the stations		77.8%	7
Other (please specify)		0.0%	0
	Total Responses		**9**

8. **How would you judge the length of time provided for completing the ratings for each of the stations? (Select ONE)**

Response	Chart	Percentage	Count
About Right		88.9%	8
Too Little Time		0.0%	0
Too Much Time		11.1%	1
	Total Responses		**9**

9. **Overall, how did you feel about participating in group discussions conducted during the ratings process for each station? (Select ONE)**

Response	Chart	Percentage	Count
Very Comfortable		77.8%	7
Somewhat Comfortable		22.2%	2
Unsure		0.0%	0
Somewhat Uncomfortable		0.0%	0
Very Uncomfortable		0.0%	0
	Total Responses		**9**

10. What level of confidence do you have that the impact data and final discussion on the final afternoon helped the panel arrive at a defensible passing score? (Select ONE)

Response	Chart	Percentage	Count
Very Confident		44.4%	4
Confident		44.4%	4
Somewhat Confident		11.1%	1
Not at All Confident		0.0%	0
	Total Responses		**9**

11. What level of confidence do you have in the final recommended passing score? (Select ONE)

Response	Chart	Percentage	Count
Very Confident		44.4%	4
Confident		55.6%	5
Somewhat Confident		0.0%	0
Not at All Confident		0.0%	0
	Total Responses		**9**

Subpanel 2

1. How clear were you about the description of the "Just Qualified" (or "Borderline Passing") candidate on the MCCQE Part II as you began the task of setting a passing score following the training on Monday morning? (Select ONE)

Response	Chart	Percentage	Count
Very Clear		30.0%	3
Clear		30.0%	3
Somewhat Clear		40.0%	4
Not Clear		0.0%	0
	Total Responses		**10**

2. Did you feel the discussion of the "Just Qualified" (or "Borderline Passing") candidate on the MCCQE Part II was helpful during the training on Monday morning? (Select ONE)

Response	Chart	Percentage	Count
Yes, Very Helpful		50.0%	5
Yes, Helpful		40.0%	4
Yes, Somewhat Helpful		10.0%	1
Not Helpful At All		0.0%	0
	Total Responses		**10**

3. **How would you judge the length of time spent (approximately 45 minutes) introducing, discussing, and editing the definition of the "Just Qualified" or "Borderline Passing" candidate? (Select ONE)**

Response	Chart	Percentage	Count
About Right		90.0%	9
Too Little Time		0.0%	0
Too Much Time		10.0%	1
	Total Responses		**10**

4. **What is your impression of the length of time for training you received for setting a passing score on the MCCQE Part II? (Select ONE)**

Response	Chart	Percentage	Count
Appropriate		50.0%	5
Somewhat appropriate		50.0%	5
Not appropriate		0.0%	0
	Total Responses		**10**

5. **How clear did you find the information that was provided regarding the scoring procedures for the MCCQE Part II? (Select ONE)**

Response	Chart	Percentage	Count
Excellent		20.0%	2
Very good		30.0%	3
Good		40.0%	4
Fair		10.0%	1
Poor		0.0%	0
	Total Responses		**10**

6. **What is your overall evaluation of the training that was provided for setting a passing score on the MCCQE Part II? (Select ONE)**

Response	Chart	Percentage	Count
Excellent		20.0%	2
Very good		50.0%	5
Good		30.0%	3
Fair		0.0%	0
Poor		0.0%	0
	Total Responses		**10**

7. **What factors influenced the ratings you made of "Just Qualified" (or "Borderline Passing") candidate responses on the MCCQE Part II? (Select ALL choices that apply)**

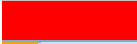| Response	Chart	Percentage	Count
The description of the "Just Qualified" or "Borderline Passing" candidate		100.0%	10
My perception of the difficulty of the stations or station components		50.0%	5
The scoring of the individual stations or station components		50.0%	5
The station statistics (e.g. candidate station scores)		20.0%	2
The statistical impact data provided before round 2		20.0%	2
Panelists discussion		20.0%	2
My experience in the field		70.0%	7
Knowledge and skills measured by the stations		60.0%	6
Other (please specify)		0.0%	0
Total Responses			**10**

8. **How would you judge the length of time provided for completing the ratings for each of the stations? (Select ONE)**

Response	Chart	Percentage	Count
About Right		80.0%	8
Too Little Time		0.0%	0
Too Much Time		20.0%	2
Total Responses			**10**

9. **Overall, how did you feel about participating in group discussions conducted during the ratings process for each station? (Select ONE)**

Response	Chart	Percentage	Count
Very Comfortable		90.0%	9
Somewhat Comfortable		0.0%	0
Unsure		10.0%	1
Somewhat Uncomfortable		0.0%	0
Very Uncomfortable		0.0%	0
Total Responses			**10**

10. What level of confidence do you have that the impact data and final discussion on the final afternoon helped the panel arrive at a defensible passing score? (Select ONE)

Response	Chart	Percentage	Count
Very Confident		50.0%	5
Confident		40.0%	4
Somewhat Confident		10.0%	1
Not At All Confident		0.0%	0
Total Responses			**10**

11. What level of confidence do you have in the final recommended passing score? (Select ONE)

Response	Chart	Percentage	Count
Very Confident		30.0%	3
Confident		40.0%	4
Somewhat Confident		30.0%	3
Not At All Confident		0.0%	0
Total Responses			**10**