MEDICAL COUNCIL OF CANADA
LE CONSEIL MÉDICAL DU CANADA

Technical report on the 2025 standard-setting exercise for the Medical Council of Canada Qualifying Examination Part I

Psychometrics and Data Science – Assessment Department

October 2025

Contents

1. Background and purpose

Standard-setting is a critical component of any high-stakes assessment program, particularly for licensing and certification decisions in the health professions. The public needs to be assured that licence and certificate holders possess the required knowledge, skills, and attitudes necessary for safe and effective patient care. Standard-setting is a process used to define an acceptable level of performance in the competency domains targeted by an examination. This level determines the standard, which is then operationalized as a numerical pass score (also known as a cut score) that is used to make classification decisions (e.g., pass or fail an exam, grant or withhold a credential, award or deny a licence).

The Medical Council of Canada Qualifying Examination (MCCQE) Part I is a computer-delivered examination that assesses the critical medical knowledge and clinical decision-making ability of a candidate at a level expected of a medical student who is completing their medical degree. Critical medical knowledge and clinical decision-making skills are assessed using multiple-choice questions (MCQs, also referred to as items).

Candidates who are graduating from medical school and who successfully complete the MCCQE Part I typically enter supervised practice. In addition to the formal accreditation processes of undergraduate and postgraduate education programs, the MCCQE Part I is the national standard for medical schools across Canada; therefore, candidates take it at the end of medical school.

Passing the MCCQE Part I is a requirement for international medical graduates applying for entry into a Canadian residency program. The exam is also used by some practice-ready assessment programs as an eligibility criterion to select internationally trained physicians for their programs.

Obtaining a pass result on the MCCQE Part I is one of the eligibility criteria that must be met to apply for the Licentiate of the Medical Council of Canada (LMCC). Canadian medical regulatory authorities may require candidates to have the LMCC to apply for a medical licence in their province or territory.

Over the past few years, the Medical Council of Canada (MCC) has been working on several grant-funded projects and seeking to improve access to its assessments and services for Canadian medical graduates, international medical graduates, and internationally educated physicians. One of these priority projects is aimed at enhancing the delivery and candidate experience of the MCCQE Part I. In addition, the MCC is working to implement improvements to the exam to ensure that results are generated and made available more quickly. Following months of research, data collection, expert consultation, and operational considerations, a new assessment model for the MCCQE Part I was defined and implementation began in April 2025. With the candidate experience being a top priority, below are some of the key adjustments:

- The overall length of the exam has been shortened; research showed that candidates reported feeling more mentally fatigued during longer exams, and while actual performance didn’t always decline, self-reported focus levels dropped (Jensen et al., 2013).
- A new break policy was introduced: breaks must be offered after 3 hours of test-taking, an improvement on the previous minimum of 4.5 hours. Candidates may accept or decline the break; those who decline can finish earlier and leave.

- The Clinical Decision-Making component of the exam, which consisted of 38 cases with short-menu and short-answer write-in questions, was removed. Critical medical knowledge and decision-making skills continue to be assessed through MCQs.
- Instead of presenting all MCQs in a single section, they are now divided into two distinct sections to provide a more balanced exam-day experience.
- Candidates are provided with more time to answer each question to help ensure the exam continues to be a fair and accurate assessment of their knowledge.

As of April 2025, the exam consists of 230 MCQs divided into two sections of 115 MCQs each. The exam includes 20 pilot questions, also called pretest items, that do not count towards candidates' total score. Candidates are allowed up to 2 hours and 40 minutes to complete each section with an optional break before the second session.

The MCCQE Part I assesses candidates using exam items mapped to two blueprint domains: Dimensions of Care and Physician Activities. This evidence-based blueprint defines the assessment framework (e.g., health promotion, acute care, diagnosis, management, communication), ensuring the exam reflects current medical practice and measures the core competencies required for safe, effective patient care in Canada.

TABLE 1.
MCCQE Part I Blueprint and exam specifications

Physician activities \ Dimensions of care	Health Promotion & Illness Prevention	Acute	Chronic	Psychosocial Aspects	Row %
Assessment/ Diagnosis					45±5
Management					35±5
Communication					10±5
Professional Behaviours					10±5
Column %	20±5	35±5	30±5	15±5	100

The MCCQE Part I is a criterion-referenced exam, meaning the pass or fail result is determined by comparing each candidate's score to the set standard (the pass score); their score is not affected by how other candidates perform. A pass means that the candidate has demonstrated the necessary knowledge, skills, and abilities as part of the requirement for medical licensure to enter supervised clinical practice in Canada. It is best practice to review the standard, and the pass score every five years (or sooner if there is a change to the exam such as implementation of a new exam format). This ensures that the standard remains appropriate and reflects the current level of competency medical graduates are expected to have to be able to practise in the profession. Not only does this

protect the public interest, it also ensures the standard reflects changes in the exam and the candidate population and stays current with advancements in medicine and medical education.

On July 15–16, 2025, a panel of 23 physicians from across Canada met at the MCC's offices in Ottawa to participate in a standard-setting exercise for the MCCQE Part I. Staff from the MCC's Assessment Department facilitated the meeting. The purpose of the meeting was to arrive at a recommended pass score for subsequent consideration and approval by MCC's Exam Oversight Committee, which is composed of physicians and medical educators from across the country and is responsible for approving the pass or fail results for candidates. This report summarizes the process, procedures, and results of the two-day exercise that led to the recommendation and approval of a new pass score for the MCCQE Part I.

2. Procedures

2.1. SELECTING A STANDARD-SETTING METHOD

Several standard-setting methods are appropriate for MCQ exams (Cizek, 2012; Cizek & Bunch, 2007). We selected the Bookmark method as our primary method based on the following considerations:

- The MCCQE Part I is a criterion-referenced exam for which a pass score should be defined as an acceptable level of knowledge and performance that candidates need to demonstrate given the intended use of the exam. A pass or fail status should be determined by comparing each candidate's performance to a performance standard regardless of the performance of other candidates. Therefore, a criterion-referenced method of standard-setting such as the Bookmark method is most appropriate for the MCCQE Part I.
- The Bookmark method is a test-centred, criterion-referenced method where expert panellists review test items and provide judgments as to the level of knowledge and performance tested by those items. Conversely, for examinee-centred approaches for performance exams, other methods for setting standards are preferred, for example, Borderline Group or Contrasting Group methods (Kane, 1998).
- The Bookmark method is a convenient way to connect a cut score to the Rasch model, which is used to calibrate items and assemble test forms for the MCCQE Part I. The Rasch model characterizes examinee ability and item difficulty simultaneously, making it possible to order items by the ability needed to have a specific probability of success and to map the items on the ability scale. In this way, candidates with scores near the location of specific items can be inferred to possess the abilities required to respond successfully to those items with the specified probability.
- The Bookmark method simplifies the cognitive complexity required of standard-setting panellists and is relatively easy to use compared to other methods.
- The Bookmark method has been used successfully for setting a standard on the MCCQE Part I in the past.

We also chose to complement the Bookmark method with the Hofstee method. The Bookmark and Hofstee methods are described below.

2.1.1. Bookmark method

The Bookmark method uses an item-mapping procedure in which one item is presented per page and items are ordered from easiest to most difficult based on operational data. Standard-setting panellists are asked to review each item in the order presented and place a bookmark where they consider a minimally competent candidate would no longer correctly answer subsequent items (see section 2.4.1.2, Defining the minimally competent candidate). De facto, this bookmark placement corresponds to the proposed cut score of each panellist.

For each item, each panellist judges whether a minimally competent candidate would have a good chance of answering the item correctly. For our purpose, we define “good chance” as having at least 0.67 probability (RP67) of answering the item correctly. In completing these judgments, panellists consider multiple factors, including (1) the knowledge being assessed by that item, (2) the difficulty level of that item and, (3) the definition of the minimally competent candidate. Panellists make this judgment for every item until they reach a point in the ordered item eBook where they feel a minimally competent candidate would no longer have a 67 per cent chance of answering items correctly. They then “bookmark” that page, which is why it is called the Bookmark method. A more detailed description of the Bookmark method is provided in *Setting Performance Standards: Foundations, Methods, and Innovations* (Cizek, 2012) and *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests* (Cizek & Bunch, 2007). It is important to note that candidates may be able to answer some items correctly beyond that bookmarked page. For example, for MCQs that have five answer choices, a candidate has a 20 per cent chance of answering an item correctly just by guessing. However, the panellists are instructed to place their bookmark where the minimally competent candidate’s chance would fall just below 0.67 probability.

Each panellist’s cut score corresponds to the Rasch ability level (RP67) associated with the bookmarked item. A cut score is derived from the median of the cut scores across panellists and then by average across subpanels. This process can be repeated in two or three rounds. Impact data (i.e., the pass or fail rates of the chosen cut score) are usually presented for discussion after each round to help panellists understand the effects of their recommendation.

2.1.2. Hofstee method

The use of criterion-referenced approaches can sometimes lead to unacceptable outcomes in the absence of political considerations associated with the decision (De Champlain, 2013). To ensure the standard set by using the Bookmark method is realistic, we also used the Hofstee method to check its reasonableness from a policy perspective. The Hofstee method is a “compromise” method that concurrently uses both a holistic judgment on an acceptable cut score (criterion-referenced) and an acceptable failure rate (norm-referenced). Panellists are asked to address four questions based on their expertise and experience in the field, their knowledge of the test content and the exam’s objective, and their understanding of the

candidate population. A cut score is derived from the panellists' answers to the following four questions:

1. What is the highest per cent pass score that would be acceptable even if every candidate attains that score?
2. What is the lowest per cent pass 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?

Panellists' answers to the first two questions provide absolute information for a criterion-referenced standard based on exam content, whereas answers to the last two questions provide relative information to define a norm-referenced standard based on candidates' performance.

The answers to each question are averaged across panellists and then plotted in a graph displaying the cumulative percentage of candidates who would fail at each point along the score scale (see section 3.3, Hofstee results). The Hofstee method is usually not used as a standalone method. For our purpose, we used it to complement the Bookmark method and provide a "reality check" on the standard set using the Bookmark method. A more detailed description of the Hofstee method is provided in *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests* (Cizek & Bunch, 2007) and in *On Educational Testing* (Hofstee, 1983).

2.2. SELECTING PANELLISTS AND ORGANIZING SUBPANELS

Selecting well-qualified panellists ensures the validity of the standard-setting process and the resulting cut score. Due to the inherent subjectivity of any standard-setting process, best practice dictates the panel should broadly represent the target candidate population with respect to background and educational characteristics (De Champlain, 2013).

In June 2024, the MCC sent out an email invitation to many physicians across the country to solicit interest in participating in our standard-setting exercise. This solicitation resulted in over 580 interested physicians, each of whom completed a demographic information sheet (see Appendix A).

Based on the demographic information provided, the MCC selected 24 participants and assigned them to two subpanels that were matched as closely as possible on key demographic variables, including (1) number of years in practice (post residency), (2) geographic region of practice, (3) gender identity, (4) medical specialty, (5) place of medical training (Canada or abroad) and (6) racial or ethnic background/identity. The main purpose of using two subpanels was to evaluate the replicability of the cut score across two parallel independent groups of physicians. This provides evidence of the replicability and generalization—and therefore the validity—of the recommended cut score. In addition, subpanels may foster more discussions as they allow each participant more opportunity to share their perspective. Although 24 participants were invited to take part in the standard-setting exercise, only 23 attended and participated. Table 2 summarizes the demographic composition of the two subpanels.

TABLE 2.
Demographic information by standard-setting subpanel

		Subpanel 1	Subpanel 2	TOTAL
Geographic region of practice	Ontario	4	4	8
	Quebec	2	1	3
	Atlantic	1	3	4
	Central	1	1	2
	Western	3	2	5
	Northern	0	0	0
	Multiple	1	0	1
Primary language of practice	French	2	1	3
	Other	10	10	20
Gender identity	Female	6	6	12
	Other	6	5	11
Racial or ethnic background/identity	Indigenous	0	0	0
	Asian	2	2	4
	Black	1	1	2
	Middle Eastern	1	3	4
	White	5	4	9
	Other	3	1	4
Medical specialty	Family Medicine	6	5	11
	Pediatrics	1	1	2
	Internal medicine	1	1	2
	Psychiatry	1	1	2
	Obstetrics	1	1	2
	Surgery	1	1	2
	Other	1	1	2
Location of work	Urban	9	8	17
	Other	3	3	6
Care setting	Hospital	4	5	9
	Other	8	6	14
Geographic location of postgraduate training	Canada	9	8	17
	Other	3	3	6
Years of practice	3–10 years	2	3	5
	11–20 years	6	6	12
	21–30 years	3	2	5
	More than 30 years	1	0	1
Total number of physicians	12	11	23	

2.3. PREPARING MATERIALS FOR THE STANDARD-SETTING EXERCISE

2.3.1. Test form selected for the ordered item eBook

Multiple test forms were used in the 2025 MCCQE Part I administration. One form was selected for the ordered item eBook that was judged to be typical in content and met strict psychometric specifications. This test form had 210 active items that were used for the standard-setting exercise. The 210 items were assembled in an eBook with one item per page, ordered from easiest to most difficult as determined by level of difficulty parameter estimates. The eBook was delivered through an online application developed by the MCC specifically for the standard-setting exercise.

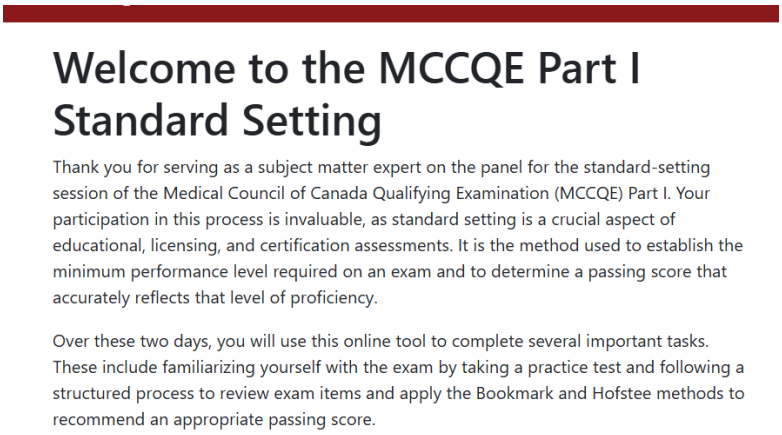


Figure 1.
Screenshot of the introductory page of the MCC's standard-setting application

2.3.2. RP67

With the Bookmark method, each panellist judges whether a minimally competent candidate has a good chance of answering each item correctly. As indicated in section 2.1.1, Bookmark method, we defined "good chance" as having at least 0.67 (RP67) probability. Though we considered other probability levels (e.g., RP50), we decided to use RP67 because it is typically used by other testing programs. RP67 is consistent with the proficiency concept used in a criterion-referenced exam, where a candidate must have a probability greater than 0.50 of answering an item correctly to meet the standard for knowledge of the assessed content. In addition, it is a relatively easy value for standard-setting panellists to understand. The ability level needed to have a 0.67 probability of answering an item correctly was calculated using this formula:

$$\theta_i=\beta_j+.708$$

where θ_i represents examinee ability and β_j represents item difficulty (Cizek & Bunch, 2007). The RP67 values were calculated for each of the 210 items selected for the standard-setting exercise.

2.3.3. Item map

An item map was prepared that included information about the order and page number of each item in the eBook, as well as each item's ID, answer key, RP67 value, and content classification.

2.3.4. Practice test and practice ordered item eBook

Panellists were provided with a practice test and a practice ordered item eBook to familiarize themselves with the MCCQE Part I and prepare themselves for the standard-setting activity. The practice test and eBook were also delivered through the MCC's standard-setting application.

The practice test consisted of 50 MCQs, representing a range of difficulty levels and content domains (Blueprint categories and discipline). The items were randomly ordered and presented to the panellists. After answering the 50 questions, they selected a button in the application to score themselves.

The practice ordered item eBook consisted of 50 MCQs on 50 pages. Items were ordered in from easiest to most difficult. Each page contained the item ID, the item, answer key, and RP67 value.

2.3.5. Background materials

Before the standard-setting exercise, panellists were given the agenda and asked to read chapter 22, "Standard setting methods in medical education," in *Understanding Medical Education: Evidence, Theory and Practice* (De Champlain, 2013). This chapter provided them with an overview of standard-setting and the Bookmark method.

2.4. ACTIVITIES DURING THE TWO-DAY SESSION

The agenda for the two-day meeting is provided in Appendix B. Day 1 was devoted to training the panellists and Day 2 was devoted to the actual standard-setting exercise.

2.4.1. Day 1 – Training and practice

Day 1 was devoted to extensive training of the panellists to ensure the credibility and defensibility of the new cut score. We began the meeting with a roundtable welcome and introduction of facilitators and panellists, as well as an overview of the purpose of the meeting. Panellists were informed that their task was to recommend a pass score—not to

make a final decision—and their recommendation would be presented to the MCC's Exam Oversight Committee for consideration and approval. We then provided an overview of the MCCQE Part I, explaining its purpose, content, format, scoring, score reporting, psychometric model, exam delivery model, and intended candidate population. We followed this with an overview of the standard-setting exercise, including its purpose and process, the selection and training of panellists, and our criterion- and norm-referenced frameworks and common methodologies. We also provided a brief explanation of the Bookmark method.

2.4.1.1. Familiarizing panellists with the MCCQE Part I

To familiarize the panellists with the type of questions and difficulty level of the MCCQE Part I, we gave them an hour to review the practice test (see section 2.3.4, Practice test and practice ordered item eBook) and answer 50 sample questions that were presented in random order. Upon completing the 50 questions, the panellists selected a button to review the items they had answered incorrectly; however, their scores were not recorded or captured by the meeting organizers or other panellists. Afterwards, the panellists discussed the difficulty level of the questions and the range of content covered while keeping in mind the purpose of the MCCQE Part I and its target candidate population.

2.4.1.2. Defining the minimally competent candidate

A critical step in any standard-setting exercise is to define the target candidate for the proficiency level targeted by the examination. For the MCCQE Part I, the target is the minimally competent candidate entering supervised practice in Canada. Before the two-day meeting, the MCC's medical education advisor and two Exam Oversight Committee members developed the definition of the minimally competent candidate, and it was approved by the Exam Oversight Committee (see Appendix C). In the afternoon on Day 1, panellists reviewed the definition and engaged in a discussion, led by the MCC's medical education advisor. Panellists were asked to envision a minimally competent candidate and discuss their characteristics, capabilities, and challenges. They were asked to think about what distinguishes a minimally competent candidate from an incompetent candidate. The intention was to help panellists converge on a unified understanding of the minimally competent candidate for the purpose of the MCCQE Part I and recommend a meaningful cut score. The discussion continued until everyone was satisfied that they had a shared understanding of the minimally competent candidate, which we asked them to keep in mind consistently throughout the standard-setting exercise.

2.4.1.3. Practising using the Bookmark and Hofstee methods

After panellists familiarized themselves with MCCQE Part I content and reached a unified understanding and definition of the minimally competent candidate, we provided step-by-step training on how to use the Bookmark method to set a cut score and reviewed the four Hofstee questions as described in section 2.1.2, Hofstee method. We divided panellists into two preorganized subpanels and assigned each panel to separate rooms with different facilitators. We gave each subpanel the chance to practise the Bookmark method using the same 50 items that were on the practice test that they completed in the morning. The only difference was that the 50 items were now ordered from easiest to most difficult. Each panellist's task was to review each item in the order presented and

provide a judgment on whether a minimally competent candidate would have at least a 0.67 probability of answering the item correctly. We asked each panellist to place their bookmark on the page beyond which a minimally competent candidate would have less than 0.67 probability of correctly answering all items (see Appendix D). During the practice, panellists were also given an item map for the practice ordered item eBook that included information about the item order, item ID, answer key, RP67 value, and content classification.

The cut score from the practice exercise was determined by the median bookmark cut scores from individual panellists on each subpanel and then the average of the two subpanels (see section 2.1.1, Bookmark method). We presented these results to the full panel for discussion, questions, and clarifications to ensure that panellists would have a good sense of the process of recommending a cut score and the impact of a cut score on pass and fail rates.

By the end of Day 1, panellists had developed a good understanding of the purpose, content, and difficulty level of the MCCQE Part I; the definition of the minimally competent candidate; the standard-setting process; and the Bookmark and Hofstee methods.

2.4.2. Day 2 – Standard-setting exercise

Day 2 started with a brief recap of Day 1 activities and a reminder to panellists of the key points about the Bookmark method. We then conducted two rounds of the standard-setting exercise.

2.4.2.1. Round 1 (preliminary round)

For round 1, we split panellists into two subpanels and placed them in separate rooms. A psychometrician facilitated each subpanel. We provided panellists with an eBook containing 210 items for standard-setting, ordered by difficulty from easiest to most difficult (see section 2.3.1, Test form selected for the ordered item eBook). We also provided an item map that included information about the item order, item ID, answer key, RP67, and content classification. We instructed panellists to review the items in order, starting from page 1, and to place a bookmark at the point beyond which they felt a minimally competent candidate would no longer have a 0.67 probability to correctly answer all items. Panellists were given three hours to provide a bookmark judgment and record their bookmark page independently (see Appendix D). During this activity, each panellist had their own printed copy of the definition of the minimally competent candidate and the purpose of the MCCQE Part I for reference.

After completing their bookmark judgments, we asked panellists to answer the four Hofstee questions as described in section 2.1.2 (see Appendix E). Specifically, we asked them to specify the highest and lowest per cent scores as well as the highest and lowest failure rates they considered to be reasonable for the MCCQE Part I. These scores were to be based on their holistic judgment, experience, knowledge of the purpose of the exam, and the definition of the minimally competent candidate.

MCC's psychometrics and data science staff collected the completed Bookmark and Hofstee data. During the panellists' lunch break, the bookmarks were tallied and the median cut scores for each subpanel were calculated. These two cut scores were then averaged to obtain the full panel recommendation. The Hofstee results were summarized at both the subpanel and full panel levels. Finally, the impact of the full panel's cut score on the failure rate was calculated using data from Canadian medical graduate first-time test-takers (CMG FTs) and all first-time test-takers (FTs) on the April 2025 MCCQE Part I.

We then presented the results and impact data from round 1 to all panellists before splitting them once again into the two subpanels in different rooms. They were given 15 minutes for discussion. Each subpanel returned to the main room with a spokesperson who brought a summary of their discussion to the full group. The full panel discussed and shared further thoughts on the process and outcomes. For comparison, panellists were also shown historical failure rates for first-time test-takers and all test-takers from 2022 to 2024, which were based on the previous cut score established in 2018.

The round 1 exercise provided panellists with realistic practice in full scale. Round 1 results, impact data, and discussions helped calibrate the panellists' understanding of the process, giving them a more unified idea about the cut score and the potential effects of their judgment. It also became clear to panellists why they needed to have a unified understanding and definition of the minimally competent candidate while judging the items. With the information learned and skills developed from round 1, panellists were better prepared for round 2—the final round.

2.4.2.2. Round 2 (final round)

In round 2, we split panellists into the same subpanels and assigned each panel to a separate room. Within each subpanel, they repeated the same exercise as in round 1. That is, they independently provided Bookmark and Hofstee judgments using the ordered item eBook for standard-setting and recorded their bookmark pages and answers to the four Hofstee questions using the forms provided. We advised that starting from the easiest item on page 1 wasn't necessary. They could target items they had been uncertain about or those located roughly 20 items before or after their Round 1 bookmark. Panellists were told they can keep the same bookmark as round 1 or change it. Again, we reminded them to keep the definition of the minimally competent candidate in mind while making judgments on the probability of answering an item correctly. They were given 1.5 hours to complete this activity.

At the end of this activity, MCC staff calculated cut scores for individual panellists, subpanels, and the full panel. They also created graphs and tables to show the impact of their cut scores on failure rates using the performance data of Canadian first-time test takers and all first-time test takers from the April 2025 administration of the MCCQE Part I.

We then presented the results and impact data from round 2 to the full panel. For comparison purposes, we again showed the historical failure rates of Canadian medical graduates who are first-time test-takers, all first-time test-takers and all test-takers from

2022 to 2024. We asked panellists to discuss the round 2 cut score, the standard-setting process used to generate it, and any potential impacts on future MCCQE Part I candidates.

2.4.2.3. Calculation of the final recommended Bookmark cut score

Each panellist's cut score corresponded to the RP67 value for the item on their bookmarked page, representing the ability required to have a 0.67 probability of a correct response as expressed on an IRT θ scale. The subpanel's cut score is the median RP67 value. The median was used instead of the mean because it is less affected by extreme values or outliers. Finally, the two cut scores from the two subpanels were averaged to obtain the full panel's cut score.

3. Results

3.1. BOOKMARK RESULTS

Table 3 presents a summary of the Bookmark cut scores for each subpanel and the full panel. Overall, there was more variability in the Bookmark cut scores in subpanel 1 than in subpanel 2; however, this difference was larger in round 1 than round 2. For each subpanel and each round, we computed the standard error of judgment, which is an estimate of the variability that we would expect if the same judging process was repeated by many different panels of similar composition. We then constructed 95 per cent confidence intervals around their cut scores using the standard error of judgment for each subpanel. In the final round, the 95 per cent confidence intervals for subpanel 1 (0.58, 0.84) and subpanel 2 (0.63, 0.90) overlapped, suggesting they did not differ significantly.

The θ cut score of 0.74 derived from round 2 became the standard-setting panel's final recommended cut score. A θ score of 0.74 translates to a scale score of 439 on the MCCQE Part I reporting scale of 300–600.

TABLE 3.
Bookmark cut scores from the MCC standard-setting exercise, July 2025

		Cut score (θ)	Minimum	Maximum	Standard deviation	Standard error of judgment
Round 1	Subpanel 1	0.70	-0.33	1.72	0.53	0.19
	Subpanel 2	0.47	0.07	1.45	0.47	0.18
	Full panel	0.58				
Round 2	Subpanel 1	0.71	0.39	0.87	0.16	0.06
	Subpanel 2	0.77	0.48	1.15	0.16	0.06
	Full panel	0.74				
Final cut score		0.74				

3.2. IMPACT DATA

As indicated earlier, we computed the impact of cut scores on failure rates using performance data from 2,734 Canadian medical graduate first-time test-takers in April 2025. These results are presented in Table 4. Round 2 had a higher failure rate than round 1. For comparison, historical failure rates are also shown for 2022, 2023, and 2024.

TABLE 4.
Recommended cut score for rounds 1 and 2 of the MCC standard-setting exercise, July 2025, with impact data and historical cut score set in 2018 and historical failure rates for reference

	Recommended cut score		Impact data	
	θ	Reported[1]	CMG* first-time test-takers	All test-takers
Round 1 (preliminary) Spring 2025	0.58	430	3.7%	22.4%
Round 2 (final) Spring 2025	0.74	439	8.3%	30.7%
	Historical cut score		Historical failure rates	
Spring 2022	0.68	226	6.9%	20.5%
Spring 2023	0.68	226	5.7%	21.2%
Spring 2024	0.68	226	5.9%	23.9%

[1]Scale of 300 to 600 with a mean of 450 and a standard deviation of 30
** CMG: Canadian medical graduate*

3.3. HOFSTEE RESULTS

Table 5 summarizes the Hofstee results computed by averaging panellists' answers to the four Hofstee questions within each subpanel and for the full panel. The results from the two rounds are very similar.

Table 5.
Summary of Hofstee results by round and subpanel

	Statistic	Subpanel 1	Subpanel 2	Full panel
Round 1	Min. acceptable percentage cut score	50	54	52
	Max. acceptable percentage cut score	68	71	69
	Min. acceptable failure rate	6	5	6
	Max. acceptable failure rate	18	13	15
Round 2	Min. acceptable percentage cut score	51	56	53
	Max. acceptable percentage cut score	68	71	70
	Min. acceptable failure rate	5	6	5
	Max. acceptable failure rate	16	13	14

Figure 1 displays the average across subpanels in round 2 plotted against a cumulative percentage of candidates who would fail at each point along the θ ability scale using the performance data of Canadian first-time test takers from the April 2025 exam session. Panellists' judgments for Hofstee indicated that the cut score should not be lower than 0.22 and not higher than 1.06. Similarly, they indicated that the failure rate should be at least 5 per cent but not higher than 14 per cent. The coordinates of maximum cut score and minimum failure rate and minimum cut score and maximum failure rate are linked by a green dotted line. The point of intersection between this line and the cumulative frequency distribution corresponds to a cut score to 0.74 if we draw a vertical line down from this point to the horizontal axis. The panellists' bookmarks were placed within these boundaries during round 2. As indicated earlier, the Hofstee method was not our primary method for setting the standard for the MCCQE Part I; it was used for a "reality check" of the standard set by using the Bookmark method. The results indicate the Bookmark cut score was consistent with panellists' global judgment of the cut score and failure rate.

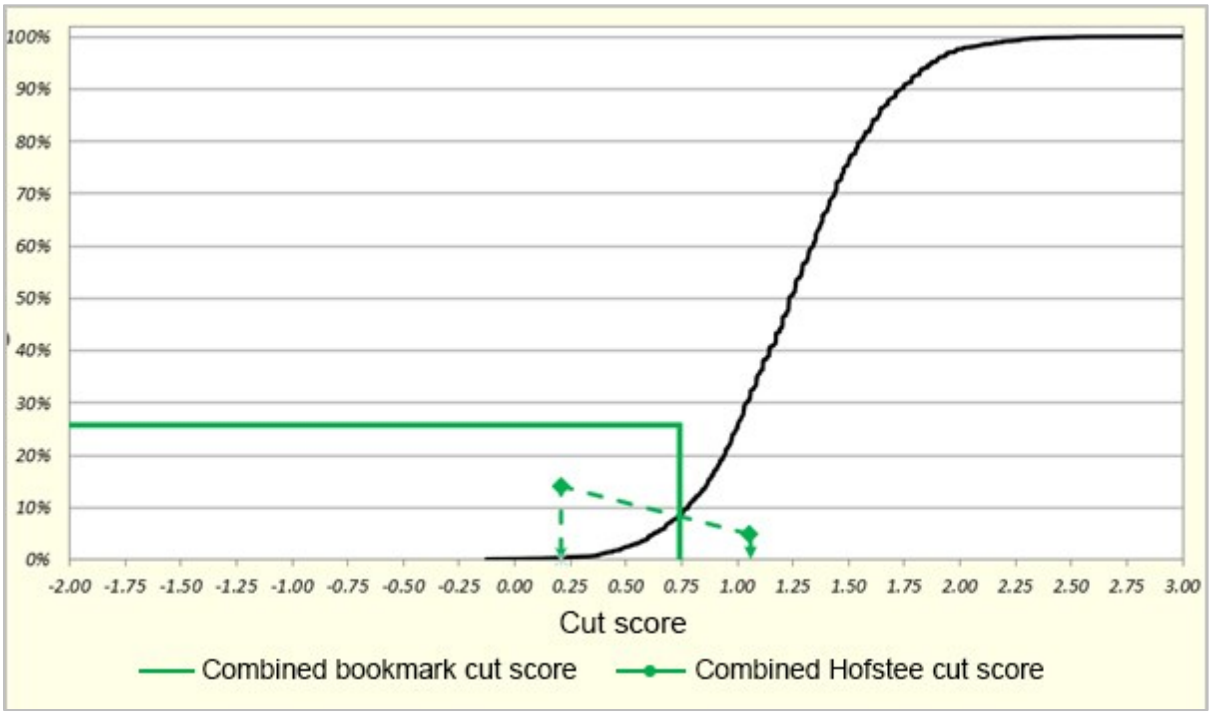


Figure 1.
Average failure rates for CMG FTs from the full panel in round 2

3.4. FINAL RECOMMENDED PASS SCORE

The panel of 23 physicians recommended a new pass score of 0.74 on the θ scale (439 on the reporting scale of 300–600) that was subsequently brought forward to the Exam Oversight Committee for consideration and approval.

3.5. POST-SESSION SURVEY

After the standard-setting exercise, panellists received a survey to provide anonymous feedback to MCC. Full results of the survey are presented in Appendix F. All 23 panellists completed the survey, although one panellist left the second half unanswered. In summary, the survey results indicate the following:

- At the beginning of the session, we provided an overview of the standard-setting process. We surveyed panellists about the clarity and quality of the information and training provided for setting a pass score: 65.2 per cent selected excellent, 30.4 per cent selected very good, and 4.4 per cent selected good.
- Central to the standard-setting exercise is the unified understanding and definition of the minimally competent candidate: 65.2 per cent were very clear about this, 26.1 per cent were clear, and 8.7 per cent were somewhat clear.
- We devoted a significant amount of time and effort to training panellists on the Bookmark procedure to ensure they knew what was expected of them before they engaged in the actual exercise. In round 1, 72.7 per cent found the Bookmark method training to be very good, 22.7 per cent found it good, and 4.6 per cent found it poor. Panellists had a better understanding of the Bookmark method in round 2, as 90.9 per cent found the information to be very good and 9.1 per cent found it to be good.
- When asked about the appropriateness of time provided for completing the standard- setting activity, 77.3 per cent thought it was about right and 22.7 per cent thought it was too much.
- At the end of each round, we presented impact data to show the effects of cut scores on failure rates and held discussions around this: 68.2 per cent found this to be very helpful and 31.8 per cent found it helpful in facilitating the panel to arrive at a defensible pass score.
- Finally, and most importantly, 50 per cent of panellists indicated they were very confident in the final recommended cut score and 50 per cent felt confident in it. None of the respondents indicated a lack of confidence.

4. Conclusions

Several findings highlight our confidence in the standard-setting process and the resulting pass score.

1. The recommended cut scores from subpanel 1 and subpanel 2 converged in round 2. This indicates that the training provided to panellists in round 1 and round 2 further reinforced their understanding of the standard-setting process.
2. The 95 per cent intervals around the cut score constructed using the standard error of judgment for subpanel 1 (0.58, 0.84) and subpanel 2 (0.63, 0.90) indicate very similar ranges and significant overlap between the two subpanels. This provides evidence to support the replicability of the cut score if different panels were used and followed the same process.
3. The recommended Bookmark cut score was within the acceptable range defined by the Hofstee method based on panellists' holistic judgment. This indicates that the criterion-referenced cut score derived using the Bookmark method is realistic and consistent with holistic judgments.
4. The results of the post-session survey indicate confidence in the training and process and procedures followed.
5. In the post-session survey, panellists expressed high confidence in the standard-setting training, the process and procedures followed, and the final recommended pass score.

These findings provide evidence to support the reliability and validity of the standard-setting process. The resulting recommended pass score is defensible from both psychometric and holistic perspectives.

The recommended pass score was presented to the Exam Oversight Committee on July 31, 2025, along with an overview of the standard-setting process and the impact data. The committee unanimously approved the recommended θ pass score of 0.74 (439 on the reporting scale of 300–600). The new pass score was implemented starting with the April 2025.

References

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). Routledge.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests* (pp. 55–189). Sage Publications.

De Champlain, A. F. (2013). Standard setting methods in medical education. In T. Swanwick (Ed.), *Understanding medical education: Evidence, theory and practice* (2nd ed., pp. 305–316). John Wiley & Sons.
(*Note*: This link is to the 3rd edition of the book.)

Hofstee, W. K. B. (1983). *The Case for Compromise in Educational Selection and Grading*. In S. B. Anderson & J. S. Helmick (Eds.), *On educational testing* (pp. 109–127). Jossey-Bass Publishers.

Jensen, J. L., Berry, D. A., & Kummer, T. A. (2013). *Investigating the Effects of Exam Length on Performance and Cognitive Fatigue*. PLOS ONE, *8*(8), e70270.

Kane, M. (1998). *Choosing Between Examinee-Centered and Test-Centered Standard-Setting Methods*. *Educational Assessment, 5*(3), 129–145.

APPENDIX A:
Demographic information sheet

MEDICAL COUNCIL OF CANADA | LE CONSEIL MÉDICAL DU CANADA

Medical Council of Canada standard-setting survey

The information requested below is being collected to help the Medical Council of Canada (MCC) select a representative pan-Canadian panel to recommend cut scores on the Medical Council of Canada Qualifying Examination (MCCQE) Part I. The standard-setting exercise is scheduled to take place in-person at MCC's headquarters, located in Ottawa, on **July 15 and 16, 2025**.

If you are interested in participating as a panellist in this activity, please submit this survey by **July 30, 2024**. Should you have any questions, contact us at *standardsetting@mcc.ca.*

Privacy statement

At the Medical Council of Canada, we are committed to protecting your privacy and ensuring the security of your personal information. The information collected through our survey will be used solely for the standard-setting exercise. All data collected through the survey, including private information such as your name, email, Licentiate of the Medical Council of Canada (LMCC) information, and telephone number, will be stored in a secure environment and accessed only by authorized personnel. We employ industry-standard security measures to prevent unauthorized access, disclosure, or alteration of your information. We do not sell, trade, or transfer your personal information to third parties. By participating in our survey, you consent to collecting, using, and storing your personal information as outlined here. If you have any questions or concerns regarding our Privacy Statement or handling your personal information, please email us at *standardsetting@mcc.ca*.

* **1. What is your full name?**

* **2. What is your preferred email address for contact?**

* **3. What is your preferred telephone number for contact?**

* **4. Do you have your Licentiate of the Medical Council of Canada (LMCC)?**

Yes

No

If yes, please provide your LMCC number:

* **5. Which of the following certifications do you have? Please select all that apply.**

Royal College of Physicians and Surgeons of Canada (RCPSC)

College of Family Physicians of Canada (CFPC)

Collège des médecins du Québec (CMQ)

None of the above

* **6. Do you have an active unrestricted licence to practise with a medical regulatory authority (MRA) in Canada?**

Yes

No

If yes, please specify which province or territory:

* **7. How many years have you been in practice since completing your residency?**

0–2 years

3–5 years

6–10 years

11–20 years

21–30 years

More than 30 years

* **8. Have you had experience supervising students/residents?**

Yes

No

* **9. How recently have you supervised students/residents?**

0–2 years ago

3–5 years ago

6–10 years ago

11–20 years ago

21–30 years ago

More than 30 years ago

* **10. Are you actively supervising students/residents?**

Yes

No

If yes, please specify how often and how many students/residents you typically supervise each year:

* **11. How many years have you been supervising Canadian medical graduates (CMGs)?**

1–5 years

6–10 years

11–20 years

21–30 years

More than 30 years

I have no experience supervising CMGs

* **12. How many years have you been supervising international medical graduates (IMGs)?**

1–5 years

6–10 years

11–20 years

21–30 years

More than 30 years

I have no experience supervising IMGs

* **13. Have you ever participated in an MCC test committee, marking session, or content development workshop?**
NOTE: Being an MCC test committee member or content development workshop participant is not a requirement to participate in the standard-setting exercise.

Yes

No

If yes, please specify the activity and when:

* **14. Have you ever been an examiner for the MCC Objective Structured Clinical Examinations (OSCEs)? Please select all that apply.**
NOTE: Being an MCC examiner is not a requirement to participate in the standard-setting exercise.

Yes, I have been an examiner for the National Assessment Collaboration (NAC) Examination.

Yes, I have been an examiner for the former Medical Council of Canada Qualifying Examination (MCCQE) Part II.

Yes, I have been an examiner for both the NAC Examination and the former MCCQE Part II.

No.

* **15. Have you participated in a candidate preparatory course / activity from a third party (i.e., not offered by the MCC) in preparation for the MCCQE Part I within the last three years?**

Yes

No

If yes, please specify the course/activity and when:

* **16. Where did you complete your postgraduate medical training?**

Canada

Other (please specify):

* **17. In what province or territory do you currently practise?**

British Columbia

Alberta

Saskatchewan

Manitoba

Ontario

Quebec

New Brunswick

Nova Scotia

Prince Edward Island

Newfoundland and Labrador

Yukon

Northwest Territories

Nunavut

Multiple provinces and/or territories (please specify):

None of the above (please explain):

* **18. What is your first language?**

English

French

Other (please specify):

* **19. What is your primary language of your medical practice?**

English

French

Other (please specify):

* **20. What is your gender identity?**

Female (including transgender women)

Male (including transgender men)

Non-binary, gender-fluid, agender, other (please specify):

Prefer not to say

*** 21. What is your race and/or ethnicity?**

African, Black, Caribbean

Central Asian

East Asian

Hispanic or Latino/Latina/Latinx

Indigenous (First Nations, Inuit, Métis)

Middle Eastern

Multiracial

South Asian

Southeast Asian

White

None of the above (please specify):

Prefer not to say

*** 22. What is your medical specialty?**

Family medicine

Internal medicine

Obstetrics and Gynecology

Pediatrics

Psychiatry

Surgery

Other (please specify):

*** 23. Which type of location best describes where you work?**

Urban

Suburban

Rural

Other (please specify):

*** 24. In which type of care setting do you primarily work?**

Hospital-based setting

Community-based setting

Other (please specify):

*** 25. I am interested in and fully available to participate in-person in the standard-setting exercise on July 14 and 15, 2025.**

Yes

No

*** 26. I am interested in participating in future standard-setting exercises.**

Yes, keep me on your list of potential participants.

No, remove my name from your list after this standard-setting exercise.

APPENDIX B:
Agenda

MCCQE Part I Standard-Setting Exercise
Maude Abbott/Thomas Roddick Meeting Rooms

JULY 15-16, 2025

DAY 1: July 15, 2025

TIME	ACTIVITIES	LEAD
08:00	BREAKFAST (and registration)	
08:30	Welcome, land acknowledgment, and introductions	
08:50	Security video and business code of conduct	
09:00	Review agenda and objectives	
09:05	Overview of MCCQE Part I	
09:25	Overview of standard-setting	
10:15	BREAK	
10:30	Panellists answer practice test (sub-panels in two rooms)	
11:45	Discussion about the practice test	
12:15	LUNCH	
13:00	Develop common understanding of the definition of the minimally competent candidate entering residency	
13:45	Training of Bookmark method	
14:15	Round 0: Split into two sub-panels and two rooms – Panellists independently practise Bookmark method using Ordered Item Booklet (ordered item eBook) for Practice Test and provide Hofstee judgments	
15:15	BREAK	
	Data entry and calculation	
15:30	Post-bookmark training discussion and clarification	
15:45	Round 0 Feedback (Practice Test): Bring sub-panels into one room – Present Round 0 results and impact data	
16:30	Wrap-up of Day 1/Overview of Day 2	
16:40	End of Day 1	
18:30	(Dinner Reservation under MCC)	

DAY 2: July 16, 2025

TIME	ACTIVITIES	LEAD
08:00	BREAKFAST	
08:30	Round 1: Split into two sub-panels and two rooms – Panellists independently provide Bookmark and Hofstee judgments using the ordered item eBook for standard-setting	
11:30	LUNCH	
	Data entry and calculation	
12:30	Round 1 Feedback: Bring sub-panels into one room – Present Round 1 results and impact data	
13:00	Sub-panels in two rooms and discuss impact data	
13:15	Sub-panels in one room for whole panel discussion	
13:30	Round 2: Sub-panels in two rooms – Panellists independently provide Bookmark and Hofstee judgments using the ordered item eBook for standard-setting	
15:00	BREAK	
	Data entry and calculation	
15:30	Round 2 Feedback: Bring sub-panels into one room – Present Round 2 results and impact data	
16:00	Complete post – standard-setting exercise survey	
16:15	Wrap-up	
16:30	End of Day 2	

APPENDIX C:
Defining the minimally competent candidate

The candidate who **attains the minimum passing score on the MCCQE Part I** can be entrusted to provide safe patient care in the Canadian context.

The candidate demonstrates the **minimum level** of foundational knowledge, sufficient clinical reasoning skills, and professionalism to reasonably assess and manage common clinical problems and identify life-threatening clinical situations.

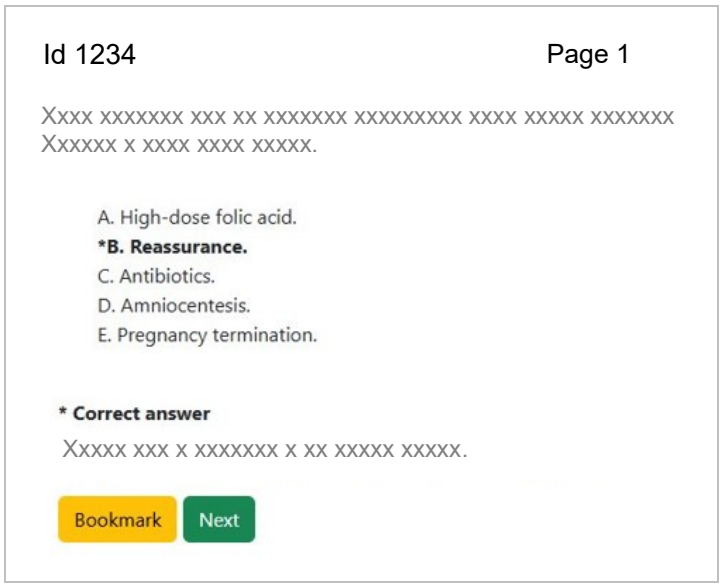
The candidate can **recognize the limits** of their own knowledge and abilities and know how to secure urgent help for patients when needed.

The candidate **is ready for the next stages** of professional development, beginning in a **supervised environment**.

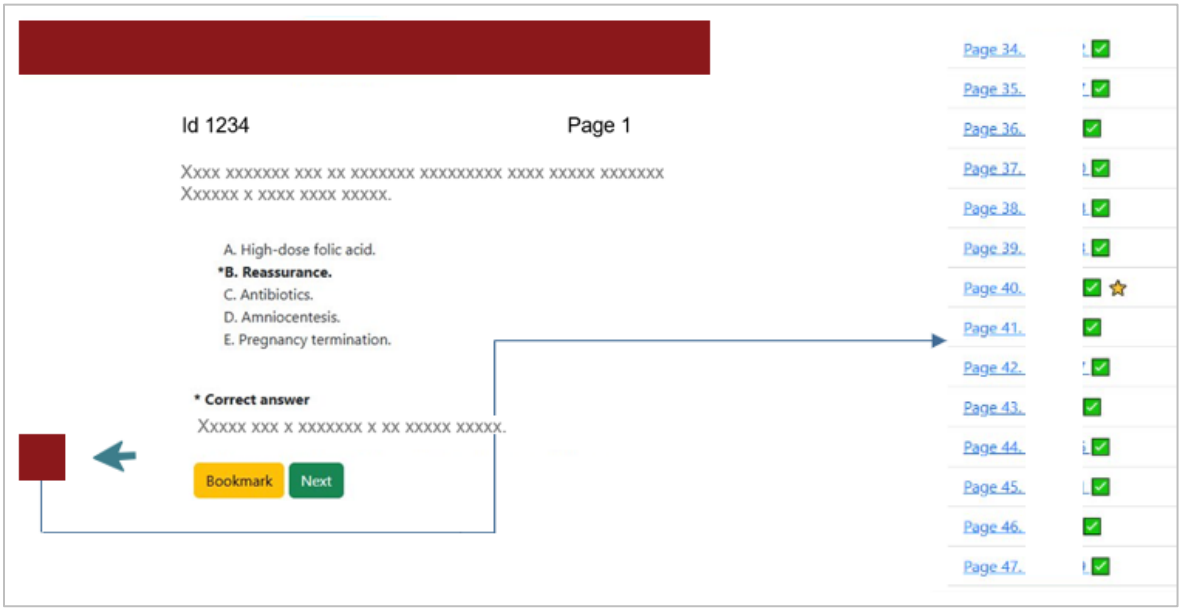
APPENDIX D:
Form to documenting the bookmark for each round

Standard-setting for the MCCQE Part I
The Bookmark method

Your task is to review the items in order until you reach a page at which, in your judgment, a minimally competent candidate would have a good chance (0.67 probability) to correctly answer all items up to that page, and their chance would decrease beyond that page. Click *Bookmark* when you reach that page.



To verify the page you have bookmarked, or the list of all items, hover over the red square on the left side.



APPENDIX E:
Form to document the Hofstee score for each round

Please answer the following four questions using integer numbers from 0 to 100%.

	Percentage
1. What is the highest percent pass score that would be acceptable, even if every candidate attains that score?	
2. What is the lowest percent pass 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 F: MCCQE Part I standard-setting 2025 – Post-session survey summary

1. Which panel were you on for the standard-setting exercise?

Response	Percentage	Count
Panel 1	52.2%	12
Panel 2	47.8%	11
Total responses	**100.0%**	**23**

2. How clear did you find the information regarding the overview of the MCCQE Part I that was provided on the morning of Day 1?

Response	Percentage	Count
Very clear	82.6%	19
Clear	13.0%	3
Somewhat clear	4.4%	1
Not clear	0.0%	0
Total responses	**100.0%**	**23**

3. During training on the morning of Day 1, did you feel the discussion about the minimally competent candidate on the MCCQE Part I was helpful?

Response	Percentage	Count
Yes, very helpful	65.2%	15
Yes, helpful	30.4%	7
Yes, somewhat helpful	4.4%	1
Not at all helpful	0.0%	0
Total responses	**100.0%**	**23**

4. Following discussion on the morning of Day 1, how clear were you about the description of the minimally competent candidate on the MCCQE Part I as you began the task of practising to set a passing score?

Response	Percentage	Count
Very clear	65.2%	15
Clear	26.1%	6
Somewhat clear	8.7%	2
Not clear	0.0%	0
Total responses	**100.0%**	**23**

5. How would you judge the length of time spent introducing and discussing the definition of the minimally competent candidate (which was around 45 minutes)?

Response	Percentage	Count
About right	95.6%	22
Not enough time	0.0%	0
Too much time	4.4%	1
Total responses	**100.0%**	**23**

6. How clear did you find the overview of standard-setting that was provided on the morning of Day 1?

Response	Percentage	Count
Very clear	56.5%	13
Clear	43.5%	10
Somewhat clear	0.0%	0
Not clear	0.0%	0
Total responses	**100.0%**	**23**

7. What is your impression of the length of training time you received for setting a passing score on the MCCQE Part I?

Response	Percentage	Count
About right	87.0%	20
Not enough time	4.3%	1
Too much time	8.7%	2
Total responses	**100.0%**	**23**

8. What is your impression of the amount of training you received for using the Bookmark method?

Response	Percentage	Count
Very adequate	65.2%	15
Adequate	34.8%	8
Somewhat adequate	0.0%	0
Not adequate	0.0%	0
Total responses	**100.0%**	**23**

9. How did you find the practice session for applying the Bookmark method on the afternoon of Day 1?

Response	Percentage	Count
Very helpful	82.6%	19
Helpful	17.4%	4
Somewhat helpful	0.0%	0
Not at all helpful	0.0%	0
Total responses	**100.0%**	**23**

10. What is your overall evaluation of the training that was provided on Day 1 for setting a passing score on the MCCQE Part I?

Response	Percentage	Count
Excellent	65.2%	15
Very good	30.4%	7
Good	4.4%	1
Fair	0.0%	0
Poor	0.0%	0
Total responses	**100.0%**	**23**

11. How could the training for setting a passing score on the MCCQE Part I have been improved on Day 1?

Overall, the feedback points towards a desire for more practice opportunities—participants felt that additional practice sessions, particularly on Day 1, would help them feel better prepared for the tests on Day 2. There's also a request for improved preparation materials, like pre-course videos or reading material to better understand concepts like *Theta*. Finally, more controlled and moderated discussions would ensure that everyone stays on track and has a chance to contribute, especially during group discussions. By balancing the time spent on explanations with more hands-on, practical exercises, and refining the session structure, the overall experience could be more effective and engaging for participants.

12. How clear were you about the definition of the minimally competent candidate for the MCCQE Part I as you began the task of setting a passing score on Day 2?

Response	Percentage	Count
Very clear	81.8%	18
Clear	18.2%	4
Somewhat clear	0.0%	0
Not clear	0.0%	0
Total responses	**100.0%**	**22**

13. How would you rate your understanding of how to apply the Bookmark Method during Round 1 of the exercise on Day 2?

Response	Percentage	Count
Very good	72.7%	16
Good	22.7%	5
Fair	4.6%	1
Poor	0.0%	0
Total responses	**100.0%**	**22**

14. How would you rate your understanding of how to apply the Bookmark Method during Round 2 of the exercise on Day 2?

Response	Percentage	Count
Very good	90.9%	20
Good	9.1%	2
Fair	0.0%	0
Poor	0.0%	0
Total responses	**100.0%**	**22**

15. What factors influenced your placement of the Bookmark on Day 2? Select all that apply.

Response	Percentage	Count
Description of the minimally competent candidate	68.2%	15
My perception of the difficulty of the test items	72.7%	16
Test item statistics (i.e., RP67)	36.4%	8
My experience with candidates in the field	59.1%	13
Knowledge and skills measured by the test items	54.6%	12
Impact data presented	50.0%	11
Panellist discussions	59.1%	13
Bookmark placement of other panellists	22.7%	5
Other (please specify)	4.6%	1
Total number of respondents	**100.0%**	**22**

16. How would you judge the length of time provided for the tasks on Day 2?

Response	Percentage	Count
About right	77.3%	17
Not enough time	0.0%	0
Too much time	22.7%	5
Total responses	**100.0%**	**22**

17. Overall, how did you feel about participating in the group discussions on Day 2?

Response	Percentage	Count
Very comfortable	77.27%	17
Somewhat comfortable	18.18%	4
Unsure	4.55%	1
Somewhat uncomfortable	0.0%	0
Very uncomfortable	0.0%	0
Total responses	**100.0%**	**22**

18. How did you find the impact data and discussions in helping the panel arrive at a passing score on Day 2?

Response	Percentage	Count
Very helpful	68.2%	15
Helpful	31.8%	7
Somewhat helpful	0.0%	0
Not at all helpful	0.0%	0
Total responses	**100.0%**	**22**

19. What level of confidence do you have in the final recommended passing score?

Response	Percentage	Count
Very confident	50.0%	11
Confident	50.0%	11
Somewhat confident	0.0%	0
Not at all confident	0.0%	0
Total responses	**100.0%**	**22**

Note: Percentages may not total 100 per cent due to rounding