



MEDICAL COUNCIL
OF CANADA

LE CONSEIL MÉDICAL
DU CANADA

Technical report on the 2026 standard-setting exercise for the Therapeutics Decision- Making Examination

**Psychometrics and Data Science –
Assessment**

August 2026

Table of contents

1. BACKGROUND AND PURPOSE.....	4
1.1. THE THERAPEUTICS DECISION-MAKING EXAMINATION.....	4
1.1.1. Types of questions on the TDM Examination	5
2. PROCEDURES	6
2.1. SELECTING A STANDARD-SETTING METHOD	6
2.1.1. Angoff method	7
2.1.2. Hofstee method	7
2.2. SELECTING PANELLISTS AND ORGANIZING SUBPANELS	8
2.3. PREPARING MATERIALS FOR THE STANDARD-SETTING EXERCISE	9
2.3.1. Test form	9
2.3.2. Background materials	10
2.4. ACTIVITIES DURING THE TWO-DAY SESSION	10
2.4.1. Day 1 – Training and practice	10
2.4.1.1. Completing a practice TDM test.....	11
2.4.1.2. Defining the minimally competent candidate	11
2.4.1.3. Practising using the Angoff and Hofstee methods	11
2.4.2. Day 2 – Standard-setting exercise	12
2.4.2.1. Round 1 (preliminary round)	12
2.4.2.2. Round 2 (final round)	13
2.4.2.3. Survey	13
3. RESULTS	13
3.1. ANGOFF RESULTS	13
3.2. IMPACT DATA.....	14
3.3. HOFSTEE RESULTS	14
3.4. FINAL RECOMMENDED PASS SCORE	16
3.5. POST-SESSION SURVEY	16
4. CONCLUSIONS	17
 REFERENCES	 18
APPENDIX A: Demographic information sheet.....	19
APPENDIX B: Agenda.....	26
APPENDIX C: Defining the minimally competent candidate.....	28
APPENDIX D: Form to document the Angoff for each round	29
APPENDIX E: Form to document the Hofstee score for each round.....	31
APPENDIX F: TDM Standard-setting 2026 – post-session survey summary	32

Tables and Figures

Table 1:	TDM Examination blueprint and specifications	5
Table 2:	Demographic information by standard-setting subpanel	9
Figure 1:	Screenshot of the introductory page of the MCC's standard-setting application	10
Table 3:	Angoff pass scores from the MCC standard-setting exercise, February 2026	14
Table 4:	Recommended pass score for rounds 1 and 2 of the MCC standard-setting exercise	14
Table 5:	Summary of Hofstee results by round and subpanel	15
Figure 2:	Average failure rates for test-takers from the full panel in round 2	15

1. Background and purpose

Standard setting is essential in high-stakes assessment programs, particularly for medical licensing and certification. The public must be confident that licence and certificate holders have the knowledge, skills, and professional judgment required for safe and effective patient care. Standard setting defines the minimum acceptable level of performance across the competencies assessed by an examination. This performance standard is then translated into a numerical pass score used to make classification decisions, such as passing or failing candidates or granting or withholding credentials.

1.1. THE THERAPEUTICS DECISION-MAKING EXAMINATION

The Therapeutics Decision-Making (TDM) Examination is a computer-delivered assessment developed and administered by the Medical Council of Canada (MCC) on behalf of Practice-Ready Assessment (PRA) programs. The examination evaluates whether candidates demonstrate the level of competence required for a family physician to practise independently and safely in Canada.

The examination assesses therapeutics, defined as the application of knowledge in the pharmacologic and nonpharmacologic management of medical conditions. Test content is organized around patient-based scenarios reflecting common and clinically significant presentations encountered in family medicine practice. These scenarios span key dimensions of care, including health promotion and illness prevention, as well as acute, chronic, and safety-related management considerations.

Items on the TDM Examination are designed to evaluate candidates' clinical decision-making and critical-thinking skills. Candidates are required to make appropriate treatment decisions, manage adverse effects and drug interactions, and provide patient-centred recommendations. The exam also assesses awareness of alternative therapies, the ability to negotiate therapeutic plans, and recognition of issues such as medication misuse and substance use disorders.

Beginning in January 2026, the TDM Examination was updated to a new format to enhance the candidate experience and operational efficiency. The revised examination consists of multiple-choice questions (MCQs) and short-menu questions (SMQs) organized into two sections. Each administration includes approximately 55 clinical cases and 110 to 140 items, with a total testing time of approximately three hours. The TDM Examination blueprint remains unchanged and consists of the Dimensions of Care and Therapeutic Interventions, as shown in Table 1.

Table 1: TDM Examination blueprint and specifications

DIMENSIONS OF CARE					
THERAPEUTIC INTERVENTIONS		Health Promotion & Illness Prevention	Acute	Chronic	Safety & Adverse Effects
	Pharmacologic				
	Nonpharmacologic				
		15%	25%	30%	30%
					70%*
					30%*
					100%

* NOTE: Health Promotion & Illness Prevention = 30% pharmacologic and 70% nonpharmacologic

The TDM Examination is a criterion-referenced assessment, in which candidate performance is evaluated against a defined standard rather than relative to the performance of other candidates. Passing is defined as demonstrating “acceptable competence.” Differences in scores among passing candidates do not imply differences in the level of acceptable competence.

It is considered best practice to review the performance standard and pass score at regular intervals (e.g., every five years) or following substantive changes to the examination, such as the implementation of a new format. Periodic review helps ensure that the standard remains appropriate, reflects current expectations for independent practice, and remains aligned with changes in the examination and the candidate population and with advances in medical practice and education.

On February 19 and 20, 2026, a panel of 20 physicians from across Canada convened at the MCC offices in Ottawa to participate in a standard-setting exercise for the TDM Examination. The session was facilitated by staff from the MCC Assessment Department. The purpose of the meeting was to establish a recommended pass score for consideration and approval by the MCC Exam Oversight Committee, a national body of physicians and medical educators responsible for approving pass–fail decisions.

This report summarizes the conceptual framework, procedures, and results of the two-day standard-setting exercise that led to the approval of a new pass score for the TDM Examination.

1.1.1. Types of questions on the TDM Examination

Short-menu questions (SMQs) are similar to traditional multiple-choice questions (MCQs), in that candidates are presented with a list of three to 25 response options. For MCQs, candidates are asked to select the single best option from the list of options presented. These questions are worth one point (dichotomously scored). For SMQs, the list of options will include two or more correct answers, and candidates will be asked to select a specified number of answers. The number of points a question is worth varies (polytomously scored).

The exam's two question types are scored using three distinct scoring methods: the single best answer method, the partial credit method, and the all-or-nothing method. In the single best answer method, a candidate can select only one option, and a score of 1 out of 1 is awarded for selecting the correct answer. In the partial credit method, a candidate receives 1 point for each correct option selected. For example, if a question has three correct options and only one is selected, the candidate receives a score of 1 out of 3. The all-or-nothing method requires a candidate to select all correct options to earn points. For example, if a question has two correct options and only one is selected, a score of 0 out of 2 is assigned; selecting both correct options results in a score of 2 out of 2.

EXAMPLES OF QUESTIONS USING ALL THREE SCORING METHODS

PATIENT-BASED SCENARIO A

A 69-year-old man presents to your clinic with a 1-year history of gradual onset of tremor in his right hand. He reports a slowing of his movements and a decreased ability to smell. He reports no other neurologic deficits nor does he report any falls. He is not taking any medications. On examination, he has a rest tremor and resistance to passive movement of his major joints. Findings of magnetic resonance imaging of the brain are normal. The patient reports that his "motor slowness" is interfering with his function.

Question 1 (example of single best answer scoring method)

Which one of the following is the most appropriate medication to prescribe? Select 1 or select "None" if none is indicated.

1. Acamprosate
2. Donepezil
3. **Levodopa ✓**
4. Quetiapine
5. Thiamine
6. None

Question 2 (example of partial credit scoring method)

Before prescribing the medication in question 1, which of the following are the most common adverse effects that should be discussed with the patient? Select 3 or select "None" if none applies.

- | | |
|------------------------|--------------------------------------|
| 1. Agranulocytosis | 8. Nausea ✓ |
| 2. Bradycardia | 9. Oral ulcers |
| 3. Diarrhea | 10. Orthostatic hypotension ✓ |
| 4. Dyskinesia ✓ | 11. QT interval prolongation |
| 5. Hepatotoxicity | 12. Seizure |
| 6. Hyperkalemia | 13. Urinary frequency |
| 7. Hypoglycemia | None |

PATIENT-BASED SCENARIO B

A 3-year-old girl is brought by her father to your clinic because he is worried about his daughter's 3-month history of lethargy. He reports that she has not been playing with other children like she used to. She has been experiencing abdominal pain from time to time. His daughter has also been fussy about food but consumes a lot of milk. She enjoys making paper airplanes that she often chews afterwards. Other than an indentation in her nails, physical examination findings are normal.

Question 1 (example of all-or-nothing scoring method)

Which of the following measures are most appropriate to address the patient's condition? Select 2.

1. Recommend more calcium-rich food
2. Encourage communication cues
3. **Recommend more iron-rich food ✓**
4. Encourage structured routine
5. Encourage interaction with other children
6. **Advise reduction of milk intake ✓**
7. Refer to a developmental pediatrician
8. Recommend a fibre-rich diet
9. Report to child protection services

2. Procedures

2.1. SELECTING A STANDARD-SETTING METHOD

In high-stakes assessments, establishing a defensible threshold for performance is a critical component of test validity. *Standard setting* refers to the process of determining this threshold in a systematic and evidence-based manner. A *pass score* defines the minimum level of performance required for candidates to be classified as passing; therefore, it is central to the interpretation and use of test results.

To establish a defensible pass score, various standard-setting methods have been developed for different contexts (Cizek, 2012; Cizek & Bunch, 2007). For criterion-referenced exams based on classical test theory, such as the TDM Examination, the most widely used approach is the Angoff method (Cizek & Bunch, 2007, p. 82). In this standard-setting exercise, we used a variation of this approach—the extended Angoff method—which is designed for exams with polytomous items.

In general, the Angoff method and its variations define the pass score based on the expected performance of a **minimally competent candidate**¹, as judged by a panel of experts. Because it relies on expert judgment, the resulting pass score may not always be aligned with practical or policy considerations. To assess its reasonableness, we also applied the Hofstee method, which provides a complementary perspective. These methods are described in the following sections.

2.1.1. Angoff method

The classical Angoff method is best suited for tests consisting of dichotomously scored items (0 or 1). In this approach, panellists estimate the probability that a minimally competent candidate would answer each item correctly.

In contrast, the extended Angoff method can be applied to polytomously scored items. For such items, panellists estimate the score (from 0 to the question's maximum score, such as 2 or 3 points; the max score a TDM question can have is 5) that a minimally competent candidate would be expected to obtain (Cizek & Bunch, 2007). In this standard-setting exercise, we had both dichotomous questions (i.e., single best answer and all-or-nothing items) and polytomous questions (i.e., partial credit items).

For each dichotomous question, panellists were asked: "Think about a group of minimally competent candidates—what percentage would answer this question correctly?"

For each polytomous question with maximum score of k (where k is the maximum number of points available), panellists were asked: "What score (0– k) would a group of minimally competent candidates be expected to obtain on this question?"

After rating all questions, each panellist's ratings were summed to obtain an individual pass score. The median pass score was then calculated for each subpanel, and the panel's recommended pass score was defined as the median of the two subpanel medians.

2.1.2. Hofstee method

The use of criterion-referenced approaches can sometimes lead to unacceptable outcomes in the absence of policy considerations associated with the decision (De Champlain, 2013). To ensure the standard set by using the Angoff 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 combines two types of judgment: an acceptable pass score and an acceptable failure rate. The first is criterion-referenced, whereas the second is norm-referenced. Panellists are asked to address four questions based on their expertise, knowledge of the test, and understanding of the candidate population. A pass score is derived from the panellists' answers to the following four questions:

¹ See Appendix C for a detailed definition.

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 plotted on an impact graph (see Figure 2). The impact graph shows the cumulative failure rate at each point along the score scale. The Hofstee method is usually not used as a standalone method. For our purpose, we used it to complement the Angoff method as a "reality check." 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 is essential to ensuring the validity of both the standard-setting process and the resulting pass score. Given the inherent subjectivity of standard setting, best practice recommends that the panel broadly reflect the target candidate population in terms of background and educational characteristics (De Champlain, 2013).

In May 2025, the MCC distributed an email invitation to physicians across Canada to solicit interest in participating in the standard-setting exercise. This call yielded over 250 respondents, each of whom completed a demographic questionnaire (see Appendix A).

Based on these data, the MCC selected 20 participants and assigned them to two subpanels. The subpanels were constructed to be comparable and representative across key demographic variables, including: (1) geographic region of practice, (2) primary language of practice, (3) gender identity, (4) racial or ethnic background/identity, (5) years in practice, (6) location of work (urban vs. rural), (7) care setting, and (8) country of postgraduate training. To achieve this balance, an evolutionary optimization algorithm was used to assign participants to the two groups.

The use of two subpanels allowed us to evaluate whether similar pass scores would be obtained across independent groups of physicians. Consistency between the two groups provides evidence for the stability and generalizability—and, by extension, the validity—of the recommended pass score. In addition, smaller subpanels may promote richer discussion by giving participants more opportunity to contribute their perspectives.

Table 2 summarizes the demographic composition of the two subpanels. Overall, the distributions of demographic variables were well balanced between the groups. Although efforts were made to include Indigenous physicians, none were available to participate.

Table 2: Demographic information by standard-setting subpanel

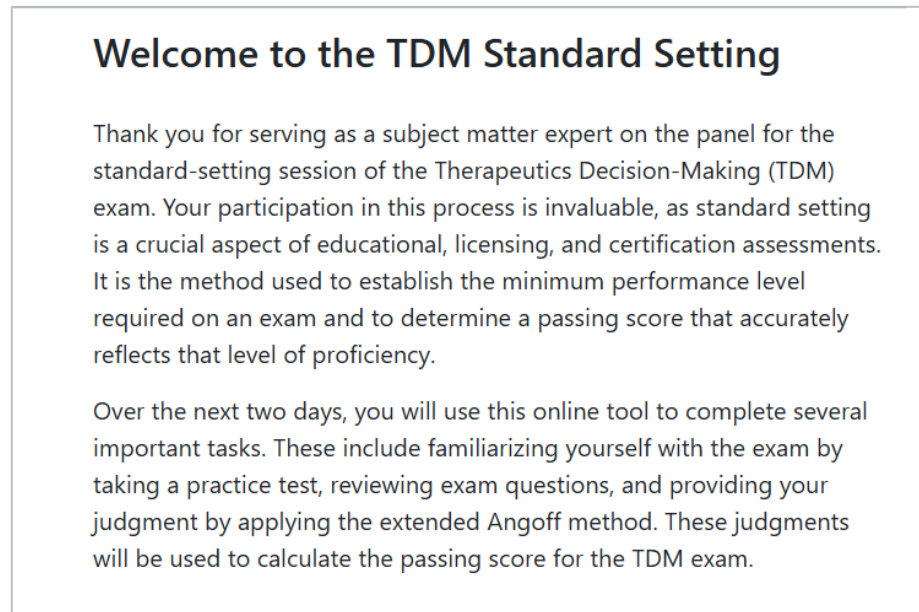
Demographic variable	Response options	Panel 1	Panel 2	Total
Geographic region of practice	Ontario	4	5	9
	Western	3	3	6
	Quebec	2	1	3
	Atlantic	0	1	1
	Northern	1	0	1
Primary language of practice	French	1	2	3
	English	9	8	17
Gender identity	Female	5	5	10
	Male	5	5	10
Racial or ethnic background/identity	Indigenous	0	0	0
	Asian	5	3	8
	Black	1	1	2
	Middle Eastern	1	2	3
	White	2	3	5
	Other	1	1	2
Years of practice	3–10 years	3	5	8
	11–30 years	6	4	10
	More than 30 years	1	1	2
Location of work	Urban	7	7	14
	Other	3	3	6
Care setting	Clinic	7	8	15
	Other	3	2	5
Geographic location of postgraduate training	Canada	8	8	16
	Other	2	2	4
Total number of physicians		10	10	20

2.3. PREPARING MATERIALS FOR THE STANDARD-SETTING EXERCISE

2.3.1. Test form

The January 2026 TDM test form was used for this standard-setting exercise. This test form had 50 active cases that were used for the standard-setting exercise. The questions were assembled in an eBook with one case per page. The eBook was delivered through an online application developed by the MCC specifically for the standard-setting exercise (see Figure 1).

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



2.3.2. 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 Angoff 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 focused on training panellists to help them become familiar with the standard-setting process. Activities included:

- Roundtable welcome and introductions of facilitators and panellists
- Overview of the meeting purpose
- Clarification of the panellists’ role (to recommend, not decide) and the approval process through the MCC’s Exam Oversight Committee

- Overview of the TDM Examination, including its purpose, content, format, scoring, reporting, psychometric model, delivery model, and target candidate population
- Practice exam
- Discussion of the minimally competent candidate
- Overview of the standard-setting process, including its purpose, procedures, panel selection and training, and the criterion- and norm-referenced frameworks used
- Brief explanation of the Angoff method
- Standard-setting practice and discussion

2.4.1.1. Completing a practice TDM test

To familiarize panellists with the format and difficulty of the TDM Examination, they were given one hour to complete a practice test consisting of 10 sample cases. After completing the cases, panellists could review items they answered incorrectly; however, their scores were not recorded or shared with organizers or other panellists.

This was followed by a group discussion on the perceived difficulty of the questions and the range of content covered, with consideration of the purpose of the TDM and its target candidate population.

2.4.1.2. Defining the minimally competent candidate

A critical step in standard setting is to define the target candidate. This candidate represents the proficiency level assessed by the exam. For the TDM, the target is the minimally competent candidate entering supervised practice in Canada. Before the two-day meeting, the MCC's senior 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 senior 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 TDM and recommend a meaningful pass 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 Angoff and Hofstee methods

After panellists familiarized themselves with the TDM content and reached a shared understanding of the minimally competent candidate, we provided them with step-by-step training in the Angoff method, along with a review of the four Hofstee questions (see section 2.1.2, Hofstee method).

Panellists were divided into two preorganized subpanels, each assigned to a separate room with a facilitator. Each subpanel then practised the Angoff method using the same 10 cases from the morning's practice test. Panellists evaluated every item and estimated the expected performance of a minimally competent candidate using the following prompts:

1. For each dichotomous question, panellists were asked: "Considering a group of minimally competent candidates, what percentage would answer this question correctly?"
2. For each polytomous question with a maximum score of k , panellists were asked: "What score (0– k) would a group of minimally competent candidates be expected to achieve on this question?"

After completing all questions, the standard-setting tool was programmed to automatically compute the sum of each item's rating as the Angoff estimate of the panellist's pass score. Once all panellists finished, we calculated the median for each subpanel, the median of the two subpanel medians (the overall pass score), and the impact data (failure rates associated with each estimated pass score). These results were verified by two analysts before being presented to the panellists for discussion.

The pass score from the practice exercise was determined by taking the median of individual Angoff estimates within each subpanel and then the median of the two subpanel medians (see section 2.1.1, Angoff method). These results were presented to the full panel for discussion and clarification to ensure that panellists understood both the process of recommending a pass score and its impact on failure rates.

By the end of Day 1, panellists had developed a clear understanding of the TDM's purpose, content, and difficulty; the definition of the minimally competent candidate; the standard-setting process; and the Angoff 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 Angoff method. We then conducted two rounds of the standard-setting exercise.

2.4.2.1. Round 1 (preliminary round)

Round 1 standard setting followed a process similar to the practice session conducted on Day 1. The key difference was that, instead of using 10 practice cases, panellists rated 50 operational exam cases. The process was as follows:

- Panellists were divided into two groups
- Each panellist independently rated all 50 cases
- Panellists completed the four Hofstee questions
- The standard-setting tool automatically calculated individual, group, and overall pass scores

- Analysts validated the computations
- Results were presented to all panellists and discussed

2.4.2.2. Round 2 (final round)

The procedure for round 2 was similar to that of round 1, except that additional information was provided. First, panellists' round 1 ratings were displayed for reference. Second, question statistics were provided, including p-values and response option selection percentages. The round 1 feedback and the additional statistics in round 2 helped panellists refine their understanding of the minimally competent candidate while considering practical constraints. As a result, round 2 pass scores tended to be more realistic.

2.4.2.3. Survey

After the standard-setting exercise was completed, panellists were asked to complete a survey about their experience. Additional details are provided in Appendix F.

3. Results

3.1. ANGOFF RESULTS

Table 3 presents a summary of the Angoff pass scores for each subpanel and the full panel. Overall, subpanel 1 set higher pass scores than subpanel 2. However, both subpanels set lower pass scores in round 2 than in round 1. 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% confidence intervals around their pass scores using the standard error of judgment for each subpanel. In the final round, the 95% confidence interval for subpanel 1 was (50.39, 62.71) and for subpanel 2 was (47.07, 54.41).

The percentage pass score of 53.65 derived from round 2 became the standard-setting panel's final recommended pass score. A percentage score of 53.65 translates to a scale score of 484 on the TDM reporting scale of 400–600².

² For the January and June 2026 TDM Examination sessions, total scores and scale scores were not reported to candidates.

Table 3: Angoff pass scores from the MCC standard-setting exercise, February 2026

		Pass score (%)	Minimum	Maximum	Standard deviation	Standard error of judgment
Round 1	Subpanel 1	71.65	59.15	79.77	6.36	2.52
	Subpanel 2	64.97	44.89	76.08	9.05	3.59
	Full panel	68.31				
Round 2	Subpanel 1	56.55	43.13	62.64	6.87	2.72
	Subpanel 2	50.74	43.93	55.42	4.10	1.62
	Full panel	53.65				
Final pass score		53.65				

3.2. IMPACT DATA

As indicated earlier, we computed the impact of pass scores on failure rates using performance data from 261 test-takers in January 2026. These results are presented in Table 4. Round 2 had a lower failure rate than round 1. For comparison, historical failure rates are also shown for 2023, 2024, and 2025. The historical failure rates for 2023–2025 include all TDM administrations in each calendar year.

Table 4: Recommended pass score for rounds 1 and 2 of the MCC standard-setting exercise

	Recommended pass score		Impact data	
	%	Reported ^a	First-time test-takers	All test-takers
Round 1 (preliminary) January 2026 ^b	68.31	521	80.4%	81.2%
Round 2 (final) January 2026	53.65	484	22.0%	22.6%
	Historical pass score		Historical failure rates	
2023	62	NA ^c	3.8%	12.5%
2024	62	NA	2.7%	10.0%
2025	62	NA	5.8%	7.2%

^a Scale of 400–600 with a mean of 500 and a standard deviation of 25.

^b These data are from the Jan. 2026 exam sessions; the standard-setting exercise took place in Feb. 2026.

^c NA means missing/unavailable.

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

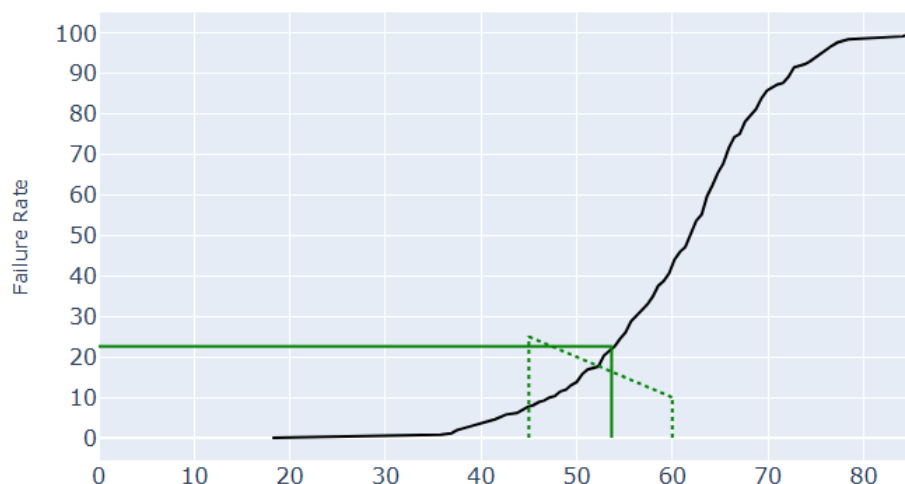
	Statistic	Subpanel 1	Subpanel 2	Full panel
Round 1	Minimum acceptable percentage pass score	52.5	60.0	55.0
	Maximum acceptable percentage pass score	70.0	70.0	70.0
	Minimum acceptable failure rate	17.5	10.0	12.5
	Maximum acceptable failure rate	30.0	30.0	30.0
Round 2	Minimum acceptable percentage pass score	47.5	45.0	45.0
	Maximum acceptable percentage pass score	60.0	60.0	60.0
	Minimum acceptable failure rate	10.0	10.0	10.0
	Maximum acceptable failure rate	30.0	25.0	25.0

Figure 2 shows the relationship between pass scores and failure rates, based on performance data from candidates in the January 2026 TDM Examination administration. The black line represents the cumulative percentage of candidates who would fail at each point along the score scale. The solid green line indicates that at the final pass score of 53.65, approximately 22.6% of candidates would fail the exam. The dotted green line represents round 2's Hofstee constraints, indicating that acceptable pass scores range from 45 to 60 and acceptable failure rates from 10% to 25%.

The coordinates representing the maximum pass score with the minimum failure rate and the minimum pass score with the maximum failure rate are connected by a green dotted line. The intersection of this line with the cumulative frequency distribution corresponds to a pass score of 52.27, as determined by projecting vertically from the intersection point to the horizontal axis.

The final Angoff pass score (53.65) in round 2 fell within these Hofstee boundaries. As noted earlier, the Hofstee method was not the primary standard-setting method for the TDM Examination but was used as a reality check on the Angoff results. Overall, the Angoff pass score was consistent with the panellists' global judgments regarding acceptable pass scores and failure rates.

Figure 2: Average failure rates for test-takers from the full panel in round 2



3.4. FINAL RECOMMENDED PASS SCORE

The panel of 20 physicians recommended a new pass score of 53.65 on the percentage score scale (484 on the reporting scale of 400–600) 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 20 panellists completed the survey. 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: 35% selected *very clear*, 65% selected *clear*.
- Central to the standard-setting exercise is the unified understanding and definition of the minimally competent candidate: 55% were *very clear* about this, 40% were *clear*, and 5% were *somewhat clear*.
- We devoted a significant amount of time and effort to training panellists on the Angoff procedure to ensure they knew what was expected of them before they engaged in the actual exercise. In round 1, 45 found the Angoff method training to be *very good*, and 55% found it *good*. Panellists had a better understanding of the Angoff method in round 2, as 85% found the information to be *very good* and 15% found it to be *good*.
- When asked about the appropriateness of time provided for completing the standard-setting activity, 90% thought it was *about right* and 10% thought it was *too much*.
- At the end of each round, we presented impact data to show the effects of pass scores on failure rates and held discussions around this: 80% found this to be *very helpful*, 15% found it *helpful*, and 5% found it *somewhat helpful* in facilitating the panel to arrive at a defensible pass score.
- Finally, and most importantly, 50% of panellists indicated they were *very confident* in the final recommended pass score, 40% felt *confident*, and 10% felt *somewhat 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 pass scores from subpanels 1 and 2 converged in round 2. This indicates that the training, impact data, and discussion after rounds 1 and 2 further reinforced their understanding of the standard-setting process.
2. The 95% intervals around the pass score constructed using the standard error of judgment for subpanel 1 (50.39, 62.71) and subpanel 2 (47.07, 54.41) indicate similar ranges between the subpanels. This provides evidence to support the replicability of the pass score if different panels were used and followed the same process.
3. The recommended Angoff pass score was within the acceptable range defined by the Hofstee method based on panellists' holistic judgment. This indicates that the criterion-referenced pass score derived using the Angoff method is realistic and consistent with holistic judgments.
4. 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 March 12, 2026, along with an overview of the standard-setting process and the impact data. The committee unanimously approved the recommended percentage pass score of 53.65 (484 on the reporting scale of 400–600). The new pass score has been applied to the TDM Examination starting with the January 2026 administration.

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. doi.org/10.4324/9780203848203
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. 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.
- 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.

APPENDIX A: Demographic information sheet



Medical Council of Canada standard-setting survey

The Medical Council of Canada (MCC) is collecting the information from this survey to help us select a representative pan-Canadian panel of **family physicians** to recommend pass scores on the Therapeutics Decision-Making (TDM) Examination. The standard-setting exercise is scheduled to take place in-person at MCC's headquarters in Ottawa on **Thursday, February 19, and Friday, February 20, 2026**.

If you are interested in participating as a panellist in this activity, please submit this survey by **May 30, 2025**. 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 this survey will be used solely for the standard-setting exercise. All data collected through this 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 the handling your personal information, email us at standardsetting@mcc.ca.

By submitting the survey, you consent to MCC collecting, using, and securely storing your information as described.

*1. What is your medical specialty?

Family medicine (includes Certificates of Added Competence)

Emergency medicine

Pediatrics

Internal medicine

Psychiatry

Obstetrics and Gynecology

Surgery

Other (please specify):

1a. Based on your previous response, it appears that you are not a family physician. Given the nature of the TDM Examination, we are prioritizing the recruitment of family physicians for the standard-setting exercise. Would you like to continue with the survey?

Yes

No

***2. Does your clinical work include a focus in any of the following areas? Please select all that apply.**

Emergency medicine

Palliative care

Addiction medicine

Other (please specify):

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

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

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

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

Yes

No

If yes, please provide your LMCC number:

***7. 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

***8. 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

Other (please specify):

***9. 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

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

Yes

No

***11. 11. 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

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

No

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

***13. 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

Less than 1 year of experience or no experience supervising CMGs

***14. 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

Less than 1 year of experience or no experience supervising IMGs

***15. 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:

***16. Have you ever been an examiner for the MCC Objective Structured Clinical Examination (OSCE)? 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.

***17. Have you participated in a candidate preparatory course or activity from a third party (i.e., not offered by the MCC) in preparation for any MCC Examination?**

No

Yes, please specify the courses or activities and when they occurred.

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

Canada

Other (please specify):

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

Alberta

British Columbia

Manitoba

New Brunswick

Newfoundland and Labrador

Northwest Territories

Nova Scotia

Nunavut

Ontario

Prince Edward Island

Québec

Saskatchewan

Yukon

Multiple provinces and/or territories (please specify):

None of the above (please explain):

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

English

French

Other (please specify):

***21. What is your primary language in your medical practice?**

English

French

Other (please specify):

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

Woman (including transgender woman)

Man (including transgender man)

Nonbinary, gender fluid, agender, or other (please specify):

Prefer not to say

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

Indigenous (First Nations, Inuit, Métis)

African, Caribbean, or Black

East Asian (e.g., Chinese, Japanese, Korean)

Southeast Asian (e.g., Thai, Vietnamese, Filipino)

South Asian (e.g., Indian, Pakistani, Bangladeshi)

Central Asian (e.g., Kazakh, Uzbek, Uyghur)

Hispanic or Latino/Latina/Latine

Middle Eastern

White

Multiracial

None of the above or other (please specify):

Prefer not to say

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

Urban

Rural

Suburban

Other (please specify):

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

Office/clinic (solo or group practice)

Hospital (hospitalist)

Hospital (emergency department)

Long-term care facility

Virtual/telemedicine

Other (please specify):

***26. I am interested in and fully available to participate in the in-person standard-setting exercise in Ottawa on Thursday, February 19, and Friday, February 20, 2026.**

Yes

No

***27. 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

TDM Standard-Setting Exercise
Maude Abbott / Thomas Roddick meeting rooms

FEBRUARY 19-20, 2026

DAY 1: February 19, 2026

TIME	ACTIVITIES	LEAD
08:00	BREAKFAST (and registration)	
08:30	Welcome, land acknowledgment, and introductions	
09:00	Security video and business code of conduct	
09:10	Review agenda and objectives	
09:20	Overview of PRA and TDM Exam	
09:45	Overview of the standard-setting process	
10:30	BREAK	
10:45	Panellists answer practice test (sub-panels in two rooms)	
11:45	Discussion about the practice test	
12:15	LUNCH	
13:00	Discussion to develop common understanding of the definition of the "Minimally Competent Candidate"	
13:45	Training of the standard setting procedure	
14:15	Round 0: Split into two sub-panels and two rooms – Panellists independently practise the standard setting procedure	
15:15	BREAK	
	Data entry and calculation	
15:45	Round 0 Feedback (Practice Test): Bring sub-panels into one room – Present Round 0 results and impact data	
16:15	Group discussion and process clarification	
16:30	Wrap-up of Day 1/Overview of Day 2	
17:00	End of Day 1	
18:00	(Dinner Reservation under MCC)	

DAY 2: February 20, 2026

TIME	ACTIVITIES	LEAD
08:00	BREAKFAST	
08:30	Round 1: Split into two sub-panels and two rooms – Panellists independently provide Angoff and Hofstee judgments using the ordered item eBook for standard-setting	
11:40	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 Angoff and Hofstee judgments using the ordered item eBook for standard-setting	
15:30	BREAK	
	Data entry and calculation	
16:00	Round 2 Feedback: Bring sub-panels into one room – Present round 2 results and impact data	
16:10	Group discussion	
16:40	Complete post – standard-setting exercise survey	
16:50	Wrap-up	
17:00	End of Day 2	

APPENDIX C: Defining the minimally competent candidate

The minimally competent candidate who just passes the TDM Examination is a candidate who is entrusted to provide **safe, effective and independent care** in the context of a Practice-Ready Assessment program.

They demonstrate the **minimum level of clinical knowledge to manage common and life-threatening conditions** using pharmacologic and non-pharmacologic approaches. They are familiar with the Canadian healthcare system, are culturally sensitive and communicate effectively. **They recognize their limits**, seek advice appropriately and refer when needed.

The minimally competent candidate may struggle with complex shared decision-making, counseling on rare conditions, and deprescribing, but **their care remains safe, appropriate and patient-centered**.

APPENDIX D: Form to document the Angoff for each round

Standard-setting for the TDM Examination The Angoff method

Instruction

Please go through each question in order. For each one, think about a group of minimally competent candidates and estimate the percentage (0–100%) of that group who would answer the question correctly.

For questions worth more than one point, indicate the score that a group of minimally competent candidates would, on average, be expected to achieve.

By clicking on “Show correct answer(s)”, a new column will appear beside the options, displaying the correct answer(s) with a “yes.”

Reminder: this process of providing your judgments will be used to calculate the pass score of the exam. There are no right or wrong percentages or scores — collectively, your judgment is essential in defining the threshold of competency required to pass the exam.

Once all panelists have completed this round, further instructions will be provided.

Partial credit item example

Question 2	
Options	Correct answer
A.	
B.	
C.	
D.	
E.	
F.	
G.	Yes
H.	Yes
I.	
What score (0–2) would a group of minimally competent candidates be expected to obtain on this question?	
1.0	
This question should be scored using the Partial credit rule.	
Max Score: 2	
Max Answers: 2	
Hide correct answer(s)	

All-or-nothing item example

Question 3

Options	Correct answer
A.	
B.	
C.	Yes
D.	Yes
E.	
F.	
G.	
H.	

Think about a group of minimally competent candidates, what percentage (0-100%) would answer this question correctly?

30

This question should be scored using the **All or nothing** rule.

Max Score: 2

Max Answers: 2

Hide correct answer(s)

APPENDIX E:

Form to document the Hofstee score for each round

Hofstee Method Round 1

In this round, you will use the Hofstee method as a “gut check” to guide your extended Angoff judgments that are used to calculate the pass score for the exam. You will be asked to answer four questions related to the acceptable ranges for the passing score and failure rate. Your answers will help create reference points that guide the final pass score for the exam.

Please answer each question using whole numbers (integers) 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: TDM standard-setting 2026 – Post-session survey summary

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

Response	Percentage	Count
Panel 1 (Maude Abbott room)	50.0%	10
Panel 2 (Thomas Roddick room)	50.0%	10
Total responses	100.0%	20

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

Response	Percentage	Count
Very clear	50.0%	10
Clear	50.0%	10
Somewhat clear	0.0%	0
Not clear	0.0%	0
Total responses	100.0%	20

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

Response	Percentage	Count
Yes, very helpful	55.0%	11
Yes, helpful	35.0%	7
Yes, somewhat helpful	10.0%	2
Not at all helpful	0.0%	0
Total responses	100.0%	20

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

Response	Percentage	Count
Very clear	15.0%	3
Clear	70.0%	14
Somewhat clear	15.0%	3
Not clear	0.0%	0
Total responses	100.0%	20

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

Response	Percentage	Count
About right	85.0%	17
Not enough time	0.0%	0
Too much time	15.0%	3
Total responses	100.0%	20

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	35.0%	7
Clear	65.0%	13
Somewhat clear	0.0%	0
Not clear	0.0%	0
Total responses	100.0%	20

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

Response	Percentage	Count
About right	100.0%	20
Not enough time	0.0%	0
Too much time	0.0%	0
Total responses	100.0%	20

8. What is your impression of the amount of training you received for using the Extended Angoff Method?

Response	Percentage	Count
Very adequate	30.0%	6
Adequate	70.0%	14
Somewhat adequate	0.0%	0
Not adequate	0.0%	0
Total responses	100.0%	20

9. How did you find the practice session for applying the Extended Angoff Method on the afternoon of Day 1?

Response	Percentage	Count
Very helpful	45.0%	9
Helpful	45.0%	9
Somewhat helpful	10.0%	2
Not at all helpful	0.0%	0
Total responses	100.0%	20

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

Response	Percentage	Count
Excellent	30.0%	6
Very good	65.0%	13
Good	5.0%	1
Fair	0.0%	0
Poor	0.0%	0
Total responses	100.0%	20

11. How could the training for setting a passing score on the TDM Examination have been improved on Day 1?

Feedback most often called for more or more representative practice questions, particularly avoiding pilot questions that differed from operational items. Other suggestions included clearer examples of how to score different question types, more context and examples when defining the minimally competent candidate, and shorter or more focused discussions. Several respondents indicated that no changes were needed.

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

Response	Percentage	Count
Very clear	55.0%	11
Clear	40.0%	8
Somewhat clear	5.0%	1
Not clear	0.0%	0
Total responses	100.0%	20

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

Response	Percentage	Count
Very good	45.0%	9
Good	55.0%	11
Fair	0.0%	0
Poor	0.0%	0
Total responses	100.0%	20

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

Response	Percentage	Count
Very good	85.0%	17
Good	15.0%	3
Fair	0.0%	0
Poor	0.0%	0
Total responses	100.0%	20

15. Which factors influenced your Extended Angoff judgment placement on Day 2? Select all that apply.

Response	Percentage	Count
Description of the minimally competent candidate	80.0%	16
My perception of the difficulty of the test items	80.0%	16
Test item statistics (i.e., % candidates selecting an option)	80.0%	16
My experience with candidates in the field	45.0%	9
Knowledge and skills measured by the test items	35.0%	7
Impact data presented	65.0%	13
Panellist discussions	60.0%	12
Ratings of other panellists	20.0%	4
Other (please specify)	10.0%	2
Total number of respondents	100.0%	20

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

Response	Percentage	Count
About right	90.0%	18
Not enough time	0.0%	0
Too much time	10.0%	2
Total responses	100.0%	20

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

Response	Percentage	Count
Very comfortable	75.0%	15
Somewhat comfortable	25.0%	5
Unsure	0.0%	0
Somewhat uncomfortable	0.0%	0
Very uncomfortable	0.0%	0
Total responses	100.0%	20

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	80.0%	16
Helpful	15.0%	3
Somewhat helpful	5.0%	1
Not at all helpful	0.0%	0
Total responses	100.0%	20

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

Response	Percentage	Count
Very confident	50.0%	10
Confident	40.0%	8
Somewhat confident	10.0%	2
Not at all confident	0.0%	0
Total responses	100.0%	20

20. Please provide any additional comments or suggestions about the setting of a passing score on the TDM Examination on Day 1 and Day 2.

Additional comments were generally positive about the organization and value of the exercise. Suggestions included providing more item-level or impact information, allowing panellists to complete the exam before rating it, clarifying the scoring of different question types, and adding a place to record item-specific comments. A few respondents also noted the need to manage discussion time and participation.

Note: Percentages may not total 100% due to rounding