Medical Council of Canada

Guidelines for the Development of Multiple-Choice Questions

February 2010

Table of Contents

PREFACE

As a member of one of the Medical Council of Canada (MCC) Test Committees or as a participant of a workshop, one of the tasks assigned to you will be the development of multiple-choice questions (MCQ) in your general area of expertise. Although this may appear as a relatively easy task at first glance, it is a skill for which we provide general guidelines to create well-constructed, reliable and valid items.

The authors hope this booklet will help guide you towards good MCQ development by reviewing basic rules pertaining to item development, provide item-writing techniques, provide examples of good and bad questions and explain our classification system.

It is my hope that you will find this guide useful.

Claire Touchie, MD, FRCPC
Vice-Chair of the MCC Central Examination Committee

Updated February 2010

Contributors

Australian Medical Council:
Ian Frank (Executive Officer)
Susan Buick (AMC Examinations Project Officer)
John Barnard (Psychometric Consultant)
Neil Paget (Technical Consultant)
Gordon Page (Project Auditor)

Australian Medical Council Test Committee Representatives:
Peter Devitt (Surgery)
Richard Doherty (Pediatrics)
Frank Hume (Psychiatry)
Barry McGrath (Internal Medicine)
Roger Pepperell (Chair BoE, OBGYN)
Neil Spike (Family Medicine)

Medical Council of Canada:
David Blackmore, (Past Director, Research & Development)
Ian Bowmer, (Executive Director, MCC)
Yves Lafortune (Manager, Evaluation Bureau)
Robert Lee (Director, Evaluation Bureau)
Jocelyne Richer (EE Test Development Officer)
Tim Wood (Manager, Research & Development)

Medical Council of Canada test Committee Representatives
Céline Bouchard (OBGYN)
David Creery (Pediatrics)
Fran Galvon (PMCH)
Sharman Robertson (Psychiatry)
Ken Taguchi (Surgery)
Donald Studney (Chair, EECC & Medicine)
Claire Touchie (Past Chair EECC, Vice-Chair CEC)
Tom Maguire (Psychometric Consultant)
Robert Charrois (Administrative Assistant)

Introduction

Multiple-choice question (MCQ) examinations have been shown to be a valid, reliable, and time and cost-effective means of assessing the cognitive skills underlying the competence of medical trainees and practitioners. This booklet has been created to aid individuals who will be developing new MCQs and reviewing existing MCQs. While writing MCQs may appear at first to be a relatively easy task, it is in fact a challenging and acquired skill that must be learned, ideally in the company of experienced question creators.

The authors hope that this booklet will serve as a guide to the development of MCQs of the highest quality. In it, the basic rules pertaining to item development will be reviewed and illustrated. Examples of good and deficient questions will be provided, illustrating many pitfalls encountered by both beginner and experienced item writers.

The Qualifying Examination Part 1 and the Evaluating Examination

Multiple-choice questions (MCQ) have long been used in the assessment of medical knowledge as they have been shown to be valid and reliable as well as time and cost-effective. Since 2000, the Medical Council of Canada (MCC) Qualifying Examination (QE) Part I has been administered in a computer-adaptive mode which allows candidates to be tested reliably using fewer questions than would be used in a traditional written examination. The application will test the candidate according to his/her ability by "adapting" to a higher or lower level of difficulty based on the candidates performance on a baseline set of questions. These groups of questions, known as testlets, include questions from different disciplines with different levels of difficulty. Therefore, not every candidate has the same set of questions during his/her examination. In turn, this means the bank of questions available must be larger than a classical examination where each candidate receives the same set of predetermined questions.

Since 2008, the Medical Council of Canada (MCC) Evaluating Examination (EE) has been administered as a computer-based examination. It is in the format of a LOFT examination (Linear-on-the-fly) which assembles the number of questions required for the examination at the start time based on predefined criteria (blueprint). As in the QE part I examination, each candidate sees a different set of questions.

Both the QE Part I and the EE use MCQs as part of their testing. All MCQs are single-correct answer (A-type). These are made up of a stem (usually a clinical scenario) with a lead-in question, followed by five response options (one correct answer and four distractors). Items to be developed should be based on the "Objectives for the Qualifying Examinations" which are available at www.mcc.ca/Objective_online/. Each item must be classified according to the classification scheme of the MCC.

Your task as a question-writer is to construct a question that will allow the candidate to demonstrate accurately what he/she knows or does not know about the objective that is being tested.

In developing MCQs for the Qualifying Examination or the Evaluating Examination, the following working guideline must be used:

"The Medical Council of Canada Qualifying Examination Part I assesses the competency of candidates after obtaining the MD degree for entry into <u>supervised</u> clinical practice in postgraduate training programs with respect to their knowledge, clinical skills and attitudes as defined by the MCC Objectives."

"The Medical Council of Canada Evaluating Examination is a general assessment of the candidate's basic medical knowledge in the principal disciplines of medicine. It is also designed to assess the skills and knowledge required at the level of a new medical graduate who is about to enter the first year of supervised postgraduate training or practice."

The Revision of Existing MCQs, and the Need for New MCQs

It is important to the MCC that a large number of high quality MCQs exists in the item banks, to permit multiple different examinations to be set with minimal use of the same questions. The MCC possesses Test Committees whose task is to create new MCQs and to evaluate the performance of existing questions. This allows poorly performing or out-of-date questions to be refined, updated, or removed from the MCQ items banks and new, up-to-date questions to be entered and tested. If needed, the MCC may also arrange MCQ item writing workshops at various locations where volunteer doctors and medical faculty gather to review existing MCQs and create new MCQs in an interactive setting.

Reviewing Existing MCQs

This section will provide some brief guidelines for reviewing existing MCQs. Existing questions are routinely reviewed to ensure that their content is correct and up-to-date. They are also reviewed in terms based on how they have "**performed**" in past examinations. The three key parameters that are used to evaluate this performance are: the item's difficulty index, the discrimination index, and the distribution of responses across its options. The review of items using these parameters is highly effective in identifying and correcting their flaws.

Item Difficulty

An item's difficulty index is the percentage or proportion of candidates that select its correct answer. The 'rule of thumb' target range for a difficulty index is **between 20% to 90%, or between .20 and .90.** It is desirable to have items with a wide range of difficulties, but preferably with a large number around the "passing" level. This allows a statistically stronger basis for identifying candidates who 'pass' or 'fail'.

Item Discrimination

An item's discrimination index is the point-biserial correlation coefficient between scores on an item (i.e., 0 or 1), and scores on the examination as-a-whole, the latter assumed to be a measure of a candidates' overall level of ability. The discrimination index can range between -1 and +1. For an item to discriminate effectively between weaker and stronger candidates, the stronger candidates should perform better on the item than the weaker candidates. If this is the case, the discrimination index will typically be in the range 0.1 to 0.4. As a general rule, an effectively discriminating item is one whose discrimination index is **positive and above 0.2.** Such items contribute to the examination's overall ability to discriminate stronger from weaker candidates.

Discrimination indices are also reported not only for the correct option, but for all options. If an item is working well, the discrimination indices for incorrect options should be lower than that for the correct option and ideally negative, indicating that the weaker candidates are selecting these options. If a discrimination index for an incorrect option is found to be higher than that of the correct option, it suggests that the stronger candidates are being attracted to that option, which may suggest a problem with the answer key, or that the question is flawed and is misleading the better candidates.

Distribution of Responses

An item's distribution of responses is a report of the percentage of candidates that select each of the five options presented in the question. As defined above, the percentage of candidates selecting the correct option defines the item's difficulty index.

If an incorrect option is selected by few or no candidates, it is not serving a useful purpose in helping identify weaker candidates. Replacements for such options are typically sought. If an incorrect option is selected by a large number of candidates, it may be suggesting that the answer key is incorrect, or that the question is in some way flawed and is misleading candidates.

The Anatomy of an MCQ

All MCQs used by the MCC are of the **single-best-answer type**. An MCQ consists of a **STEM**, a **LEAD-IN**, and five **OPTIONS**, one of which is the keyed or correct response, four of which are **DISTRACTORS.**

Stem

The stem is a short description of a **clinical scenario** of a common or a clinically important patient presentation. It should be clear and include all the information necessary for the candidate to reason out the clinical problem. These data may include:

- Age, Gender (e.g., a 45-year-old man)
- Site of Care (e.g., comes to the Emergency Department - only if needed to answer the question)
- Presenting Complaint (e.g., because of headache)
- Duration (e.g., that has continued for 2 days).
- Patient History
- Physical Findings
- +/- Results of Diagnostic Studies
- +/- Initial Treatment, Subsequent Findings, etc.

How much data you provide in the stem is determined by what the question is testing. If the question is testing an aspect of a data acquisition related to history taking, physical examination, or investigations, the stem is often brief. If the question is testing a data interpretation skill (e.g., making a diagnosis) or a management skill (e.g., treatment), more extensive data is often given to provide a basis for the clinical decision being tested.

The stem must pose a clear question, and it should be possible to arrive at an answer without reading the options. To determine if the question is clear and focused, cover up the list of options with your hand -- if the question is clear, the good candidate should be able to answer it without reading the options.

An example of a **good stem** for a moderately difficult and complex problem is seen below:

A 58-year-old man presents with sudden onset of left-sided chest pain associated with shortness of breath, palpitations, and dizziness. His past history is relevant for a recent diagnosis of a lung carcinoma. Which one of the following historical facts is most useful in establishing the etiology of his diagnosis?

Why is this a **good stem**?

1. All essential features (age, gender) are given.
2. The underlying condition of this scenario (pulmonary embolism) is important as failure to diagnose and treat correctly could be fatal.
3. It is relatively more common when a malignancy is present.
4. It is terse; it can be read and assimilated quickly yet is clinically rich.
5. It lends itself to the asking of a number of additional clinically important questions such as how to confirm the diagnosis, how to treat, associated features to look for, etc.
6. The author can make up three or four different questions using the same stem.

Lead-In

The **lead-in** is the question being asked and should be the last sentence in the stem.

In the case above, the lead-in is: "Which one of the following historical facts is most useful in establishing the etiology of his diagnosis?" Here, the candidate is being tested on his/her ability to reason that this probably is a pulmonary embolism case, and that there may be other findings that would strengthen this. For example, lower limb swelling would suggest a deep vein thrombosis possibility brought on by a hypercoagulable state of carcinoma of the lung.

Another lead-in for the same stem could be: "Which one of the following tests is the most likely to confirm the diagnosis?" Here, the candidate would need to select from a number of plausible diagnostic tests. Note that the lead-in will often stress selecting the option that is the 'most likely' or the 'best' answer -- a test of clinical judgment.

Options (Possible Answers)

These are the **five** options which represent possible answers to the question. One of the options should be the correct answer (the "**key**"). The other distractors, may be plausible but not the best choice. Distractors need to be constructed with great care. Some rules are:

- All incorrect options or distractors should be homogeneous with each other and with the correct answer. They should fall into the same category as the correct answer (e.g.,: all diagnoses, tests, treatments, prognoses, or disposition alternatives). All distractors should be plausible, grammatically consistent, logically compatible, and relatively the same length as the correct answer.
- Each distractor should be plausible and none should stand out as being obviously incorrect. Common misconceptions and faulty reasoning provide a good source of plausible options, as do the mistakes that are often made by a minimally competent candidate. You should be able to provide the line of reasoning that a candidate would use to select any one of the distractors as an answer. If you are unable to provide the line of reasoning then more than likely, the distractor is not plausible.
- Do not use "All of the above" or "None of the above" as options.

Failure to offer distractors that are homogenous may allow an lower performing candidate to deduce the correct answer by analyzing the pattern of distractor inhomogeneity rather than by using medical knowledge. Examples of faulty distractors are found in some of the example MCQs which follow.

Let us look at the “good” item again, this time focusing on the **options**.

A 58-year-old man presents to the Emergency Department with sudden onset of left-sided chest pain associated with shortness of breath, palpitations, and dizziness. His past history is relevant for a recent diagnosis of a lung carcinoma. Which one of the following historical facts is most useful in establishing the etiology of his diagnosis?

1. Daily recurrent fevers.
2. Purulent sputum.
3. History of hypercholesterolemia.
* 4. Swollen tender lower extremity.
5. Radiation of pain to left arm.

Note that most but not all of the incorrect options are “findings”, which are elements of history or physical findings that our candidate might expect to use at the bedside. All are plausibly associated with chest pain and possibly lung carcinoma. Option #1 would attract the candidates thinking of metastatic disease or pneumonia. Option #2 would also pull in those thinking of pneumonia. Option # 3 is a poor option. The Test Committee will probably change it. “History of hypercholesterolemia is a “risk factor” and could lead to heart disease causing chest pain and shortness of breath. However it is different enough from 1, 2 and 5 that an examination-wise candidate could eliminate this option because of its inhomogeneity. The question thus becomes easier than intended.

If the question was to be used in this original format, the statistics would probably show that a disproportionately low percentage (e.g.,1%-3%) of candidates would have selected #3 as the key. The option would be deemed to have “performed poorly”. A better option in this case would be, “Presence of an enlarged lymph node in the neck”, which one might reasonably find in a lung cancer case, might assume a metastatic disease (pain and shortness of breath). Option #5 would draw candidates focusing on the emergency room setting, the chest pain, and palpitations. Option #4 “Swollen tender lower extremity” would be selected by candidates who focused on the etiology as requested and not the diagnosis per se. They would need to think of the diagnosis of pulmonary embolus because of the sudden of onset of shortness of breath and to connect the lung cancer with the risk of hypercoagulability and thus the risk of deep vein thrombosis and a swollen lower extremity.

Formats and Styles to Avoid

The "**EXCEPT" Format**

Avoid creating questions that ask the candidate to select a wrong answer from among four correct ones. Such questions are most often a test of factual recall. In addition, such questions can be very distracting when they appear in a cohort of normal A-type questions and can unfairly punish an otherwise good candidate.

All of the following statements regarding obstructive jaundice are correct EXCEPT?

1. It is associated with an elevated conjugated bilirubin.
2. It is often associated with impacted gallstones.
* 3. It is not caused by external compression.
4. Diagnosis may require an ERCP.
5. Laboratory abnormalities usually include an elevated alkaline phosphatase.

Other things wrong with this question:
1. No clinical stem.
2. Cannot answer question if options covered.
3. Use of a negative option (#3) in an **EXCEPT** question is logically/grammatically difficult.
4. Use of acronym (ERCP) without writing it out in full.
5. Heterogeneity of options with etiology, laboratory, and diagnostic/therapeutic tests.
6. Use of vague terms 'usually', 'often', and 'may'.
7. The item tests factual recall only.

THE "FACTOID" Format

This form of question simply asks a candidate to recall a single fact. This format **is not acceptable and** will not be used.

The axillary nerve supplies which one of the following muscles?

* 1. Deltoid.
2. Supraspinatus.
3. Teres major.
4. Infraspinatus.
5 Latissimus dorsi.

A much more acceptable question testing this same theme or knowledge would be:

You are examining a young man who dislocated his shoulder 1 month ago and now complains of muscle weakness and loss of sensation. He is likely to have difficulty with which one of the following actions?

1. Extending his arms behind his back.
2. Scratching his shoulder blades from below.
3. Pressing his hand down on the table.
* 4. Raising his arm above his head.
5. Using a key to open a door.

The positive features of this question are:
1. Clinical stem is present.
2. Clinically relevant that the axillary injury is often missed when treating a dislocation.
3. Candidate could *reason* answer by basic knowledge of anatomy, recall of clinical teaching, or actual experience in Emergency.
4. Writer could make another good question regarding the zone of sensory loss.

FACTOID questions typically identify a diagnosis or a clinical problem in the question stem (e.g., "For a patient with a herniated disc in the cervical spine…"), and ask for the clinical features of that problem, how to treat the problem, etc. Any candidate who has memorized the textbook chapter on the signs, symptoms, and treatments associated with the clinical problem should get the item correct through a process of factual recall. In contrast, if an item requires a candidate to reach a conclusion, solve a problem, or select a treatment, the question will be testing the application of knowledge. These types of questions will usually consist of a clinical vignette with patients presenting with a set of signs and symptoms, and the candidate must make decisions about what additional information to gather (history, physical, laboratory); what the data they already possess means; how and where to manage the patient, etc.

Generating Additional Questions Using the Same Stem

You can create more than one question using the same clinical stem, but targeting different aspects such as physical examination, diagnostic tests, establishing a diagnosis, treatment, prognosis, complications, education, and or risk factors.

A 58-year-old man presents to the Emergency Department with sudden onset of left-sided chest pain associated with shortness of breath, palpitations, and dizziness. His past history is relevant for a recent diagnosis of a lung carcinoma. Which one of the following historical facts is most useful in establishing the etiology of his diagnosis?

1. Daily recurrent fevers.
2. Purulent sputum.
3. History of hypercholesterolemia.
* 4. Swollen tender lower extremity.
5. Radiation of pain to left arm.

Questions testing the ordering of investigations:

A 58-year-old man presents to the Emergency Department with sudden onset of left-sided chest pain associated with shortness of breath, palpitations, and dizziness. His past history is relevant for a recent diagnosis of a lung carcinoma. Which one of the following tests is most likely to confirm his diagnosis?

1. Electrocardiography.
2. Chest radiography.
* 3. Ventilation-perfusion scan.
4. Echocardiography.
5. Holter monitoring.

A 78-year-old man presents to the Emergency Department with sudden onset of left-sided chest pain radiating to his back. He has a past history of stable angina and peripheral vascular disease. His blood pressure is 80/50 mmHg with a heart rate of 120/minute. Which one of the following tests would most likely confirm the diagnosis?

1. Electrocardiography.
2. Chest radiography.
* 3. Computerized tomography of the chest.
4. Echocardiography.
5 Ventilation-perfusion of the lung.

As a general rule no more than three or four questions should be made from one stem as this makes databank management difficult.

Item Cloning

Item cloning is most often the process of developing questions which 'look like' existing questions, but which in fact are different questions with different correct answers. For example, you can develop a question portraying a patient with same demographics, clinical setting, and presenting complaint as in an existing question, but whose subsequent clinical data are somewhat different, suggesting a different underlying cause, the need for different investigations, the need for different treatments, etc. These questions can present the same option list as the question being cloned, but the answer will be different.

Another approach to cloning is to develop questions which test the same 'objective' as an existing question, but do so with a new patient with a different age or gender, perhaps in a different clinical setting, but maintaining the same clinical information (presenting complaint, history, and physical data, etc.) in the stem and presenting the same list of options.

Developing item clones is often an easier task than developing totally new items, and cloned items make it much more difficult for examinees to develop 'black market' copies of the examination questions. A problem with cloned items is that they are often similar in what they test, and thus cannot be used on the same examination.

Examples of Well and Poorly Constructed MCQs

The following examples will illustrate both well and poorly-constructed questions. Potential problems with the questions are pointed out.

Example 1 – well constructed

A 76-year-old man is brought to the Emergency Department by relatives who state that he had collapsed suddenly but regained consciousness within minutes. There was no seizure activity. His electrocardiogram showed a sinus rhythm (76/minute), a right bundle branch block, and left anterior fascicular block (left axis deviation). Which one of the following is the most likely cause for this man's loss of consciousness?

1. Ventricular tachycardia.
2. Type I second degree atrioventricular block (Wenckebach).
3. Paroxysmal supraventricular tachycardia.
* 4. Intermittent heart block.
5. Atrial flutter with 2:1 atrioventricular block.

Comments:

1 Well-constructed item – based on a clinical vignette, tests data interpretation skills.
2 All necessary information provided.
3 Correct answer and options are homogeneous and plausible.

Example 2 – flawed

An 86-year-old woman fell at the local nursing home and sustained an intertrochanteric fracture of her left hip. On clinical examination, you would expect to find her left leg

1. shortened, abducted and internally rotated.
2. lengthened, abducted and internally rotated.
3. shortened, adducted and externally rotated.
* 4. shortened, abducted and externally rotated.
5. lengthened, abducted and externally rotated.

Comments:

1. You would not be able to answer this item correctly if the options were covered.
2. Although the options are homogeneous, candidates can use the convergence strategy. That is, shortened, abducted, and externally rotated are the most commonly used variables of all the options and therefore may be the most logical answer to someone who has little content knowledge.
3. The item is a test of factual recall, as in essence it is asking, "What left leg clinical findings are associated with an intertrochanteric fracture of a left hip?"
4. A better way to construct this item would be to put the clinical findings in the stem and to ask what type of fracture the patient possesses.

Example 3 – flawed

A 24-year-old woman presents to a walk-in clinic with fever, flank pain, frequency, and dysuria. The urinalysis (urine microscopy) shows 1+ proteinuria, 25 white blood cells per high power field, and a few granular casts. Which one of the following investigations is the next best step?

1. Intravenous pyelography.
2. Intravenous antibiotics.
3. Creatinine clearance.
* 4. Midstream urine culture.
5. Oral analgesia.

Comments:

1. Well-constructed stem.
2. Lead-in question asks for investigations; therefore, distractors 2 and 5 can be eliminated as they are treatments; this is an example of a logical cue.
3. Could either change distractors 2 and 5 to be investigations, or change the lead-in question to "Which one of the following is the next best step in the management of this patient?"

Example 4 -- flawed

You see a 45-year-old woman because of a sudden loss of consciousness. On examination, her vital signs are normal, she is not pale, and she is not diaphoretic. Which one of the following is more typical of "fainting" as a conversion symptom than of a syncopal attack due to orthostatic hypotension?

1. Bradycardia.
2. Muscle twitching.
* 3. Absence of pallor and sweating.
4. Urinary incontinence.
5. Rapid recovery.

Comments:

1. Stem contains elements that are directly related to the correct answer. Both lack of pallor and sweating (diaphoresis) are mentioned in the stem which leads to the correct answer. This is an example of word repeats.
2. The clinical information (the case) is actually not necessary; all that is needed is the final sentence.
3. The question does not test clinical decision-making; rather it tests factual recall.

The following examples of two well constructed questions demonstrate how to make an item easier or harder:

Example 5

A 62-year-old man presents with a few days' history of peripheral edema and decreased urine output. On examination, his blood pressure is 195/90 mmHg with 3+ pitting edema of his lower extremities. His creatinine is 230 μmol/L (70-120) and urinalysis shows 2+ leukocyte esterase with 3+ proteinuria. Which one of the following is the most likely diagnosis?

1. Urinary tract infection.
2. Urolithiasis.
* 3. Nephrotic syndrome.
4. Hepatorenal syndrome.
5. Congestive heart failure.

Example 6

A 62-year-old man presents with a few days' history of peripheral edema and decreased urine output. On examination, his blood pressure is 195/90 mmHg with 3+ pitting edema of his lower extremities. His creatinine is 230 μmol/L (70-120) and urinalysis shows 2+ leukocyte esterase with 3+ proteinuria. Which one of the following is the most likely diagnosis?

1. Hypertensive nephropathy.
2. Chronic pyelonephritis.
* 3. Nephrotic syndrome.
4. IgA nephropathy.
5. Allergic interstitial nephritis.

Comment:
Using the same stem, you can make an item easier or more difficult by varying the distractors. As a guide to deciding how easy or difficult to make a question, think about the level of performance you would expect from the candidates being tested by this examination – a level equivalent to that expected of Canadian medical school graduates entering supervised practice.

Other Important Question Writing Principles

1. Refer to drugs using their generic name. In the event that you feel that candidates will be unfamiliar with the generic drug name, the trade name should be inserted in parentheses.
2. American English spelling will be used for all questions. e.g., "Hemoglobin" instead of "haemoglobin".
3. If needed, the use of acronyms should follow the full spelling of terms; e.g., "magnetic resonance imaging (MRI.)".
4. The "*Clinical Laboratory Tests - Normal Values*" (appended) are available as a reference to the candidates by clicking on the appropriate icon on the computer-based examination.
    a. EE
        i. All of the normal values are provided in the stem, in parentheses.
        ii. More than three laboratory values in a stem should be in table format with an asterisk identifying abnormal values.
    b. Part 1
        i. More than three laboratory values in a stem should be provided in table format.
        ii. Any laboratory reference value, which is not listed in the "*Clinical Laboratory Tests - Normal Values*" page, should be inserted in parentheses following the result recorded in the question. As well, all normal values for pediatric questions must be provided in the question, in parentheses.
5. Indicate the correct response to your questions with an asterisk.
6. Options should be numbered 1 to 5.
7. Provide a reference for each test item that you write. The reference should include the author's name, title of text, edition, publisher, and page number.
8. Images can be incorporated in items as a vehicle for testing important clinical data interpretation skills. Their use is encouraged. They must be digitalized. They must NOT be copyrighted and/or must be accompanied by a patient consent if provided from your personal collection.

Checklist for the Development of Test Items

The following lists summarize the guidelines for writing effective MCQs.

Guidelines for Item Structure

The basic steps to follow in the development of your test items are:

1. Using the MCC Objectives (www.mcc.ca/Objectives_Online/), select a clinical problem/presentation of relevance to the skill level of a recent medical school graduate, and think about the key steps that should be taken in resolving the problem. Depending on the problem selected, these steps can relate to taking a history, examining the patient, ordering or interpreting investigations, recognizing a need for urgent action prior to a full work-up, defining the diagnosis, treating, counseling, follow-up, prevention, etc. The question you develop should test an important clinical decision relative to one of these key steps.
2. Develop or select a case that you will present in the stem of the question. It is a good strategy to base the case on a real patient.
3. Write the stem of the question.
4. Formulate the correct or best answer.
5. Develop four additional plausible answers (for a total of five possible answers).

Guidelines for Item Stems

1. The stem should be written in a clear and concise manner. Clinical vignettes should be used whenever possible and should include all necessary information which is relevant to the case at hand. The stem must include a lead-in question that poses the clinical challenge being presented to the candidate (e.g., identify the most likely diagnosis, identify the most important next step in management).
2. The question format for the lead-in should be, in most cases, a direct and complete question. It must not be an incomplete-sentence to be completed by selecting one of the options.
3. A well-constructed stem should contain all the necessary content for a competent candidate to answer the item without having to read any of the options. A good test of this rule is to cover the options with your hand and try to choose the correct answer.
4. Do not use the negative form of question (‘EXCEPT” questions).
5. Avoid “tricky” and overly complex items. The goal is to test the candidate’s competency in medicine, not to confuse or trick them.

Guidelines for Correct Answers

1. The correct answer should be clearly correct and defensibly better than the other options. Your lead-in question will often ask, What is the best/most likely/most important/…?
2. Avoid making the correct answer clearly longer/shorter than the other options.
3. Avoid clues to the correct answer such as:
   a. Using textbook wording in the correct answer and not in the options.
   b. Using specific inclusive or exclusive determiners such as *always, never,* or vague terms such as *seldom, frequently*, etc.
   c. Using words in the correct answer that are also used in the stem.
   d. There being no link between the stem and some of the distractors (e.g., stem asks for investigations, distractors include treatments).
   e. There being a lack of parallelism among the options (grammatical, structural, vocabulary, technical jargon).

Guidelines for Distractors

1. Distractors should be consistent with the stem.
2. All distractors should be plausible. The distractors should include content which is reflective of common misconceptions or errors. If the correct answer is of the *best*-type, can competent candidates make a case for one or several options? If so, reformulate those options.
3. Avoid the use of humor as this can potentially be distracting to candidates.
4. Never use *"All of the above"*, *"None of the above".*
5. Acceptable distractors
   a. Are homogeneous in content (e.g. all are diagnoses or all are therapies).
   b. Are incorrect or definitely inferior to the correct answer.
   c. Do not contain any hints to the correct answer.
   d. Would seem plausible and attractive to the uninformed.
   e. Are similar to the correct answer in construction and length.
   f. Are not mutually exclusive to each other or to information in the stem.
6. Examples of irrelevant difficulty
   a. Options are long, complicated or double.
   b. Numerical data is not stated consistently.
   c. Terms in the options are vague (e.g., rarely, usually).
   d. Language in the options is not parallel.
   e. Options are in a non-logical order.

Summary

As stated in the introduction to this guide, writing MCQs is a challenging and an acquired skill. The guidelines, offered on the preceding pages, outline the essence of that skill. Writing MCQs can also be an intellectually rewarding process, and many question authors remark that it is a remarkable form of continuing professional development (CPD), especially when the process of writing and editing questions is pursued interactively with colleagues. For those who may wish to go beyond the core guidelines in this document, an excellent source of more detailed guidelines on writing MCQs is: Case, S. M., Swanson, D. B., "Constructing Written Test Questions for the Basic and Clinical Sciences," National Board of Examiners (Philadelphia, 2002). This book is available online, and at no cost, from the website www.nbme.org.

The MCC would welcome your comments on this guide and your suggestions for its improvement. Please send your comments to Tanya Bennett, Test Development Officer, at tbennett@mcc.ca or to Réjeanne Nasrallah, Production Assistant, at rnasrallah@mcc.ca, both at the Medical Council of Canada.

Appendix A

Specifying Item Classifications (for the MCCQE Part 1 Examination)

MCQs are classified relative to blueprint parameters, and as part of the item development process, you will be asked to classify your items using these parameters. For each parameter, the item is assigned a classification number which identifies a specific area of interest. These classifications provide information to help in the selection of items for a particular examination, and also assist in identifying areas in which additional items need to be written for the item bank. The following list identifies the item specification parameters. (Indicate here where more information is found on the breakdown within each of these parameters.)

MCC Item Classification for MCCQE Part 1 Examination

A: MCC Objective category
- available at www.mcc.ca/Objectives_Online/

B: Diagnostic category (ICD-10)
- available at http://apps.who.int/classifications/apps/icd/icd10online/

D: Discipline
- Medicine
- Obstetrics & Gynecology
- Pediatrics
- PHELO
- Psychiatry
- Surgery

I: Clinical Task
- I01 Obtain history
- I02 Obtain physical signs
- I03 Obtain laboratory data
- I04 Mechanisms of disease / etiology / pathogenesis
- I05 Interpret data / make a diagnosis
- I06 Management non drug therapy
- I07 Management drug therapy
- I08 Management education / counseling
- I09 Management follow-up / compliance
- I10 Procedure
- I11 Prognosis / Complications / Outcomes
- I12 Prevention
- I13 Communication

Sample Questions with Classifications

A 32-year-old alcoholic man, who had a mastoidectomy as a child, presents with headaches, nausea, vomiting, drowsiness, and confusion. He is afebrile. On examination, you note that his right eardrum is not visualized, there appears to be some discharge, and there is slight neck stiffness. Which one of the following is the most appropriate investigation at this time?

1. Lumbar puncture.
2. Electroencephalography.
3. Skull radiograph.
* 4. Computerized tomography scan of the head.
5. Blood culture.

A. MCC Objective: 028 Ear Pain
B. Diagnostic: G01 Meningitis in Bacterial disease classified elsewhere
D: Discipline: Medicine
I. Clinical Task: I03

An 84-year-old woman presents with a history of confusion and constipation. Laboratory investigations reveal:

Serum calcium	**2.9 mmol/L**
Creatinine	**146 μmol/L**
Hemoglobin	**108 g/L**

Which one of the following is the most likely diagnosis?

1. Hyperparathyroidism.
2. Chronic renal failure.
* 3. Multiple myeloma.
4. Vitamin D intoxication.
5. Renal cell carcinoma.

A. MCC Objective: A58-2 Delirium/Confusion
B. Diagnostic: C90.0 Multiple Myeloma
D. Discipline: Medicine
I. Clinical Task: I05

A 32-year-old woman presents with a 2-week history of diarrhea associated with heat intolerance, sweating and restlessness. Physical examination reveals a blood pressure of 150/60 mmHg and a pulse of 106/minute. She has a fine tremor of her outstretched arms. Her thyroid is diffusely enlarged, firm, and tender. Which one of the following tests will help to establish the etiology of her problem?

1. Antithyroid antibodies.
2. Sensitive thyroid-stimulating hormone assay.
3. Free triiodothyronine (T_3.)
* 4. Radioactive iodine uptake.
5. Erythrocyte sedimentation rate.

A. MCC Objective: A63 Neck Mass/Goiter/Thryoid Disease
B. Diagnostic: E05 Thyrotoxicosis
D. Discipline: Medicine
I. Clinical Task: I04

Appendix B

Item Classifications for the MCC Evaluating Examination

MCQs are classified according to the categories of the master blueprint. These classifications provide information to help in the selection of items for an examination, and assist in identifying areas in which additional items need to be written for the item bank.

Each category of the blueprint is assigned a classification code. The classification of items is part of the item development process. You will be asked to classify each new item by assigning the corresponding classification codes.

The following list identifies the blueprint categories:

A. MCC Objective
B. Discipline
C. Clinician Task
D. Patient Group

Sample Questions with Classifications

A 32-year-old alcoholic man, who had a mastoidectomy as a child, presents with headaches, nausea, vomiting, drowsiness, and confusion. He is afebrile. On examination, you note that his right eardrum is not visualized, there appears to be some discharge, and some slight neck stiffness. Which one of the following investigations is the most appropriate at this time?

1. Lumbar puncture.
2. Electroencephalography.
3. Skull X-ray.
* 4. Computed tomography scan of the head.
5. Blood culture.

A.	MCC Objective	A58-1	Coma
B.	Discipline:	D028	Medicine
C.	Clinician Task:	CT001	Data-gathering and interpretation
D.	Patient Group:	HG003	Adult Health

An 84-year-old woman presents with a history of confusion and constipation. Laboratory investigations reveal:

Serum calcium	*****	**2.9 mmol/L**	**2.18-2.58**
Creatinine	*****	**146 μmol/L**	**50-90**
Hemoglobin	*****	**108 g/L**	**123-157**

Which one of the following is the most likely diagnosis?

1. Hyperparathyroidism.
2. Chronic renal failure.
* 3. Multiple myeloma.
4. Vitamin D intoxication.
5. Renal cell carcinoma.

A.	MCC Objective	A58-2 Delirium/Confusion
B.	Discipline:	D028 Medicine
C.	Clinician Task:	CT001 Data gathering and interpretation
D.	Patient Group:	HG003 Adult Health

A 32-year-old woman presents with a 2-week history of diarrhea associated with heat intolerance, sweating, and restlessness. Physical examination reveals a blood pressure of 150/60 mmHg and a pulse of 106/minute. She has a fine tremor of her outstretched arms. Her thyroid is diffusely enlarged, firm and tender. Which one of the following tests will help to establish the etiology of her problem?

1. Antithyroid antibodies.
2. Sensitive thyroid-stimulating hormone assay.
3. Free triiodothyronine (T_3.)
* 4. Radioactive iodine uptake.
5. Erythrocyte sedimentation rate.

A.	MCC Objective	A63 Neck mass/Goiter/Thyroid Disease
B.	Discipline:	D028 Medicine
C.	Clinician Task:	CT001 Data-gathering and interpretation
D.	Patient Group:	HG003 Adult Health

A 65-year-old man, who had been making an apparently satisfactory recovery from a myocardial infarction six days previously, suddenly develops pulmonary edema. There is a regular tachycardia of 120/minute, a parasternal heave, a pansystolic murmur over the precordium, and a S3 gallop. Blood pressure is 100/60 mmHg. Which one of the following is the most likely diagnosis?

1. Post-infarct pericarditis.
2. Another myocardial infarction.
* 3. Ruptured papillary muscle.
4. Cardiac tamponade.
5. Ventricular aneurysm.

A.	MCC Objective	A27	Dyspnea
B.	Discipline:	D028	Medicine
C.	Clinician Task:	CT002	Data gathering and interpretation
D.	Patient Group:	HG003	Adult Health

Appendix C

CLINICAL LABORATORY TESTS - NORMAL VALUES

This table lists reference values for the most common laboratory tests and is intended for interpretation of the results as they are provided in the examination. Note that all values are provided in SI units. All values apply to adults.

Many important laboratory reference values are not listed here, because of the less frequent use of these tests. Such values are inserted in parentheses following the result recorded in the examination questions.

BLOOD

COAGULATION (HEMOSTASIS)	
Bleeding time (Ivy)	< 9 minutes
International Normalized Ratio (INR)	0.9-1.2
Partial thromboplastin time (PTT)	28-38 seconds
Prothrombin time (PT)	10-13 seconds
HEMOGRAM	
Hematocrit (Hct)	
Female	0.370-0.460
Male	0.380-0.500
Hemoglobin (Hb)	
Female	123-157 g/L
Male	130-170 g/L
Mean corpuscular volume (MCV)	80-100 fL
Mean corpuscular hemoglobin (MCH)	27-34 pg
Platelet count	130-400 X 10^9/L
Red blood cells (RBC)	
Female	4.0-5.2 X 10^{12}/L
Male	4.4-5.7 X 10^{12} /L
Red cell distribution width (RDW)	11.5-14.5%
Reticulocyte count	20-84 X 10^9/L
Erythrocyte Sedimentation rate (Westergren)	
Female	< 10 mm/hour
Male	< 6 mm/hour
WHITE BLOOD CELLS & DIFFERENTIAL	
White blood cell count (WBC)	4-10 X 10^9/L
Segmented neutrophils	2-7 X 10^9/L
Band neutrophils	< 0.7 X 10^9/L
Basophils	< 0.10 X 10^9/L
Eosinophils	< 0.45 X 10^9/L
Lymphocytes	1.0-4.0 X 10^9/L
Monocytes	0.1-1.0 X 10^9/L
CHEMICAL CONSTITUENTS	
Albumin (serum)	35-50 g/L
Alkaline phosphatase (serum)	38-126 U/L
Aminotransferase (transaminase) (serum)	
Alanine (ALT; SGPT)	17-63 U/L
Aspartate (AST; SGOT)	18-40 U/L
Gamma glutamyl transferase	
Female	10-30 U/L
Male	10-48 U/L
Ammonia (plasma)	9-33 μmol/L
Amylase (serum)	< 160 U/L
Bicarbonate (HCO_3) (serum)	24-30 mmol/L
Bilirubin (serum)	
Direct (conjugated)	< 7 μmol/L
Total	< 26 μmol/L
Calcium (serum)	
Total	2.18-2.58 mmol/L
Ionized	1.05-1.30 mmol/L
Chloride (serum)	98-106 mmol/L
Cholesterol (serum)	< 5.2 mmol/L
Low density lipoprotein (LDL)	< 3.37 mmol/L for low risk
Low density lipoprotein (LDL)	< 2.0 mmol/L for high risk
High density lipoprotein (HDL)	> 0.9 mmol/L
Cortisol (serum)	160-810 mmol/L
Creatine kinase (CK) (serum)	20-215 U/L
Creatinine (serum)	
Female	50-90 μmol/L
Male	70-120 μmol/L
Ferritin	
Female	11-307 μg/L
Male	24-336 μg/L
Folic (Folate)	> 15 nmol/L
Glucose fasting (serum)	3.3-5.8 mmol/L
Hemoglobin A_{1C}	4-6%
Iron (serum)	11-32μmol/L
Lactate dehydrogenase (LDH) (serum)	95-195 U/L
Lipase (serum)	< 160 U/L
Magnesium (serum)	0.75-0.95 mmol/L
Osmolality (serum)	280-300 mmol/kg
Oxygen saturation (arterial blood) (S_aO_2)	96-100%
P_aCO_2 (arterial blood)	35-45 mmHg
P_aO_2 (arterial blood)	85-105 mmHg
pH	7.35-7.45
Phosphorus (inorganic)(serum)	0.8-1.5 mmol/L
Potassium (K) (serum)	3.5-5.0 mmol/L
Prostate Specific Antigen (PSA)	0-4 μg/L
Protein (serum)	
Total	60-80 g/L
Albumin	35-50 g/L
Sodium (Na) (serum)	135-145 mmol/L
Thyroid-stimulating hormone (sensitive)	0.4-5.0 mU/L
T3 (free)	3.5-6.5 pmol/L
T4 (free)	8.5-15.2 pmol/L
Total Iron Binding Capacity (TIBC)	45-82 μmol/L
Transaminase - see Aminotransferase	
Triglycerides (serum)	< 2.20 mmol/L
Troponin T (TnT)	< 0.01μg/L
Urea nitrogen (BUN) (serum)	2.5-8.0 mmol/L
Uric acid (serum)	180-420 μmol/L
Vitamin B_{12}	133-674 pmol/L

CEREBROSPINAL FLUID

Cell count	< 10 x 10^6/L
Glucose	2-4 mmol/L
Proteins (total)	0.20-0.45 g/L

URINE

Calcium	< 7.3 mmol/day
Chloride	110-250 mmol/day
Creatinine	6.2-17.7 mmol/day
Osmolality	100-1200 mOsm/Kg
Potassium	25-120 mmol/day
Protein	< 0.15 g/day
Sodium	25-260 mmol/day