GUIDELINES FOR THE DEVELOPMENT OF KEY FEATURE PROBLEMS & TEST CASES

August 2012 (v3)

Table of Contents

PREFACE

Dear reader-user:

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 or review of key features problems in your general area of expertise for the clinical decision making (CDM) section of the MCC Qualifying Examination (MCCQE) Part I. Although this may appear as a daunting task, it is a skill for which the MCC provides general guidelines to create well-constructed, reliable and valid items for the CDM component of the MCCQE Part I.

This booklet will help guide you towards developing good CDM problems and test cases by reviewing basic rules pertaining to key features, clinical scenarios, write-in and short-menu questions, and scoring keys.

I would like to gratefully express my appreciation for the thorough review and editorial comments provided by Drs. Gordon Page and Georges Bordage.

I trust that you will find this guide useful. If you have any comments or suggestions for improving this set of guidelines, please forward them to the MCC CDM Test Committee.

Pierre Plourde, MD, FRCPC
Chair, MCC CDM Examination Committee

August, 2012

I.   INTRODUCTION

The key features (KF) test format, first discussed at the Cambridge Conference in 1984, was originally included in the MCC's Qualifying Examination (MCCQE) Part I in 1992 to replace Patient Management Problems (PMPs) and to avoid a sole reliance on multiple-choice questions as the basis for judging qualification for licensure (Bordage & Page, 1987; Page & Bordage, 1995; Page, Bordage & Allen, 1995). Key features problems have subsequently been adopted by other groups, including medical schools for in-house examinations and colleges of physicians for use in certification and fellowship examinations.

The key features approach is based on the notion of "case specificity", namely that clinical performance on one problem is not a good predictor of performance on other problems (Elstein, Shulman & Sprafka, 1978). Typically, less than 10 to 15% of the variance in performance in one problem is shared with any other problem. Each problem contains unique challenges. Consequently, assessments of clinical performance need to sample broadly because skills do not generalize across problems. In order to sample broadly in a fixed amount of time (3-4 hours), assessment is best served by focusing exclusively on the unique challenges, that is, the key features, in the resolution of each problem, be they essential issues or specific difficulties; the test then contains many short, focused clinical cases.

When the key features approach was implemented in 1992, four main benefits were envisioned:

1.   Broader sampling of cases with increased reliability compared to PMPs;
2.   More focused assessment of case-specific clinical decisions, that is, the key features;
3.   Varied response formats to accommodate question types compared to the unique latent-image format of PMPs; and
4.   Defensible pass-fail decisions.

The development of key features cases for the Clinical Decision Making (CDM) part of the MCCQE Part I has been guided by psychometric considerations of content validity and test score reliability, and by sound principles of test development. Key features cases provide a great deal of flexibility on issues of question format, multiple responses to questions, and scoring criteria. Key features problems have been found to be useful in the assessment of clinical decision-making skills that require not only medical knowledge but the ability to apply that knowledge within clinical scenarios requiring critical decisions at specific decision points during the assessment and management of the problem. These specific critical decision points are what constitute the "key features" of the problem.

Since November 2000, the MCCQE Part I is a computer-based examination. In the change from a paper-and-pencil examination, the clinical decision-making component was transformed into a multi-form examination whereby examinees are administered one of several parallel forms, each form containing a certain number of cases. A statistical equating design was chosen to construct the groupings of clinical cases based on content as well as psychometric criteria. Clinical cases are grouped in sets of cases called caselets. Caselets are comprised of 6 clinical cases covering the following major specialty areas of medicine: medicine, obstetrics-gynecology, pediatrics, psychiatry, surgery, and population health, ethical, legal, and organizational aspects of medicine.

In developing key features problems for the MCCQE Part I, the following considerations must be kept in mind when preparing test items:

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

On an examination composed of key features cases, the following is an example of a typical test case and questions that a candidate would see.

A.   *Example of a Key Features Test Case and Questions.*

At 20:00h, the nurse in the emergency department asks you to see a 24-year-old woman, who has 2 children and is now 31 weeks pregnant, reporting of bright red bleeding coming from her vagina that began about two hours ago. The three sanitary napkins that she used were completely soaked. Her pregnancy has been normal, as were the two previous ones. She has not had any pain in her belly and she says that the fetus is moving as usual.

Her BP is 110/70 mm Hg and her pulse is 92/min. Your examination of the abdomen reveals a uterine height of 31 cm with a soft, non-tender uterus. The fetus is in a breech position and has a heart rate of 150/min. No bleeding has occurred since she arrived 25 minutes ago.

Question 1. What is your leading diagnosis at this time? List only one or write "normal" if you judge this situation to be within normal limits.

1. _____

Question 2. What steps will you take in your immediate assessment and management of this patient? Select as many as are appropriate.

1.   Artificial rupture of membranes
2.   Cervical swab for chlamydia
3.   Complete blood count
4.   Computed tomographic scan of abdomen & pelvis
5.   Cross and match for 2 units of blood
6.   Discharge home with instructions to come back if bleeding starts again
7.   Fetal scalp monitor to assess status of fetus
8.   Human immunodeficiency virus serology
9.   International normalized ratio (INR)
10.   Large bore intravenous line
11.   Partial thromboplastin time (PTT)
12.   Pelvic ultrasound
13.   Pelvic examination
14.   Syphilis serology
15.   Ultrasound using vaginal probe
16.   Vaginal swab for Group B streptococcus

The preparation of key features test materials is based on three steps:

1.   Selecting problems from a domain, in this case, the MCC Objectives;
2.   Defining key features for each problem, given a specific age group and setting; and
3.   Writing test cases and questions, as well as setting the scoring keys for each key feature tested.

The template used to prepare the test case and questions for the example above is illustrated below.

Example of a Key Features Template for the Third Trimester Bleeding Problem.

Clinical problem	Third Trimester Bleeding
MCC objective	Vaginal Bleeding (112); Pregnancy - Obstetrical Complications (80-3)
Age group	Adult
Clinical situation	Single typical problem - life-threatening event
Site/setting	Emergency department

Key features: Given a pregnant woman experiencing third-trimester vaginal bleeding with no abdominal pain, the candidate will:

1. Consider placenta previa as a leading diagnosis;
2. Avoid performing a pelvic examination (because it may cause fatal bleeding);
3. Avoid discharging home from the emergency department; and
4. Order pelvic ultrasound (to confirm placenta location).

Scoring key: Each key feature receives 1 point; key feature points are averaged for each case.

Question 1   KF-1

Score	*Criteria*	*Synonyms*
1.0	Placenta previa	Marginal placenta, low placenta, low insertion
0	Listing more than one response	

Question 2   KF-2

Score	*Criteria*
1.0	Did NOT select #13. Pelvic examination
0	#1. Artificial rupture of membranes
	#2. Cervical swab for chlamydia
	#9. Insert fetal scalp monitor to assess status of fetus
	#13. Pelvic examination
	Selecting more than six items

Question 2   KF-3

Score	*Criteria*
1.0	Did NOT select #6.
	Discharge home with instructions to come back if bleeding starts again
0	#6. Discharge home with instructions to come back if bleeding starts again
	Selecting more than six items

Question 2   KF-4

Score	*Criteria*
1.0	#12. Pelvic ultrasound
0	#15. Ultrasound using vaginal probe
	Selecting more than six items

II. TESTING CLINICAL DECISION MAKING VERSUS KNOWLEDGE

Key features cases are designed to specifically assess decision-making skills rather than simple recall of factual knowledge. While knowledge is obviously a very important requisite for effective problem solving, the challenge posed by key features cases is the application of knowledge to the resolution of a problem — that is, the use of knowledge to guide decisions to elicit clinical clues, to formulate diagnostic impressions, to order investigative or follow-up procedures, to acquire data to monitor a course of action or evaluate the severity or probability of an outcome, or to select a management course. For example, key features cases should not assess examinees' ability "to describe features of delirium tremens" (a knowledge issue); rather, they should assess their ability "to recognize delirium tremens in a specific patient" (a clinical reasoning issue) and "to prescribe appropriate therapeutic measures" (a clinical decision issue). The distinction is not just semantic but focuses on different skills. While it may be easier to list the cardinal features of delirium tremens, it may be altogether different and more challenging to recognize delirium tremens within a clinical scenario and to prescribe an appropriate course of action.

The line between testing knowledge and testing clinical decision making may become thin when a particular key feature hinges on the possession of a single piece of factual knowledge. For example, the key feature "Prescribe the appropriate dose of diazepam for a prolonged convulsion (status epilepticus)", is an example of this situation. If this clinical action is recognized as a critical step in the resolution of the problem of prolonged convulsion, that is, a key feature for this problem, then it may be tested using a key features format.

The following example provides an illustration of questions that test only recall of knowledge versus clinical decisions and actions. Such "recall of knowledge" questions should not be used in key features cases.

A 20-year-old nulliparous type I diabetic woman presents to your office to discuss the complications of diabetes in pregnancy.

1.   ***What are the maternal complications of type I diabetes in pregnancy? List up to four.***

　　1.　　_____
　　2.　　_____
　　3.　　_____
　　4.　　_____

2.   ***What are the fetal complications of type I diabetes in pregnancy? List up to two.***

　　1.　　_____
　　2.　　_____

The concerns with such questions are that they only test the recall of factual information and that they could in fact be presented independently of the case. That is, the clinical scenario is not necessary to answer the questions. To improve this situation, the case could be re-written to present a poorly controlled diabetic pregnancy and the first question could then ask, "Which investigations would you order at this time?". A general rule to keep in mind when developing key features cases is that if the question asked can be answered without reference to the attached clinical scenario, then it is not a good question and is likely not measuring clinical decision making.

III.   THE KEY FEATURES CONCEPT

The term "key features" was introduced by Bordage and Page following a review and analysis by Norman et al (1985) of the research on the nature and assessment of CDM skills. A key feature is defined as:

1.   a critical or essential step(s) in the resolution of a problem,
2.   a step(s) in which examinees (in our case, graduating medical students) are most likely to make errors
3.   in the resolution of the problem, or
4.   a difficult or challenging aspect in the identification and management of the problem in practice.

The definition of a key feature highlights the fact that not all steps in the resolution of a problem are equally important and that testing time is better spent by focusing on evaluating the critical or challenging steps, the key features. The key features represent discriminating features between successful and unsuccessful candidates. Consequently, the use of key features to develop test cases will result in more discriminating measures of competence.

The concept of key features represents two important shifts from traditional positions in the assessment of clinical competence. First, it shifts emphasis from the methods of assessment to the object of assessment. The first question to be addressed in developing a key features problem is, "What is the problem to be assessed?"; selection of the problem must be guided by and directly linked to the MCCQE objectives. Once the MCCQE objective-linked problem has been selected, the next question to tackle is, "What are the critical, essential, or challenging elements in the resolution of this problem?," that is, "What are this problem's key features that should be assessed?". Then, and only then, the subsequent question is, "Which method or methods are best suited to measure the key features for this particular problem?" Thus, the object of assessment is clearly linked to the MCCQE objectives and the methods are adapted to the object.

Second, key features shift the emphasis from assessing all aspects of solving a problem to assessing only the essential elements relative to each problem. This shift recognizes that the essential elements in resolving a problem are unique or specific to each problem (referred to as "case specificity"). For some problems, the key features may pertain to data gathering or data interpretation while for others they may focus on choosing an appropriate therapeutic or follow-up plan. As a newcomer to preparing key feature cases and questions, the shift from testing thoroughness to only testing selected aspects of the clinical decision making process, the key features, is one of the most challenging aspects of the test development task. For example, the key feature may only focus on one diagnosis within a differential diagnosis rather than the entire set of diagnoses (e.g., in a case of knee pain, the key feature may focus only on including septic arthritis in the differential diagnosis because that is the diagnosis most likely to be missed by incoming unsupervised residents). From a key features' perspective, only septic arthritis will be given points; the other elements of the differential diagnosis will not be credited with any points. If one views septic arthritis as the signal, that is, the crucial element of the differential diagnosis to assess, it will stand out better by only rewarding its inclusion in the differential diagnosis, rather than drowning it in noise (the other elements of the differential diagnosis). By focusing exclusively on critical decisions, in this case thinking about a septic arthritis, the discriminating power of the question, and the exam, is greatly increased, that is, the question (exam) better discriminates between strong and weak candidates. The signal-to-noise ratio is maximized.

Key features are often unique to different cases or presentations of a clinical problem, as they may vary relative to the clinical presentation of the problem and relative to other issues such as the patient's age and gender. It is therefore unusual to have a "generic" set of key features for a specific clinical problem. For example, the critical elements in the resolution of a dia¬betic problem as a life-

threatening event (e.g., coma) are quite different from those of a diabetic problem presenting as an undifferentiated complaint (e.g., fatigue in an adolescent). Similarly, the critical elements in resolving a problem of respiratory failure in an 18-year-old man may differ from those for a simi¬lar problem in a 70-year-old woman.

The key features approach is appealing for assessing CDM skills. The clinical problems are relatively short and focus only on a few critical elements in the resolution of the problem (i.e., the key features), permitting more problems to be included on an examination within a fixed period of time (e.g., 30-40 key feature cases compared to 10-12 PMPs in a half-day exam). This broader sampling of problems has direct implications for improving the reliability of examination scores. The key features problem format also permits a flexible approach to question format, the number of options to choose from, and instructions regarding the number of allowable responses. More specifically, the short-answer question format is available for situations where cueing from a list of options would compromise measurement of the decision-making process of the examinee, or where listing options would provide too great a crutch to weaker examinees, the individuals to whom the entire examination process is most often directed. Finally, the scoring keys are flexible in terms of the number and configuration of correct responses to a question and can accommodate the complexity and configurations of actions often required in the resolution of clinical problems.

A focus on critical steps and broad sampling of problems (36 cases with respect to the MCCQE Part I) provide a foundation for the content validity of a key features examination format. That is, if such an examination is constructed from a carefully developed blueprint, it will consist of a representative and adequate sample of clinical problems from the domain of problems for which examinees are responsible and the questions within each problem will test only the important steps in its resolution—each problem's key features. Flexibility in formats and scoring keys emphasizes the strengths of this case-based examination format in accommodating the realities and complexities of clinical medicine.

IV.   PREPARING KEY FEATURES EXAMINATIONS

As indicated previously, the preparation of key features test material occurs in three steps:

1.   Selecting problems from the domain, in this case, the MCC Objectives,
2.   Defining key features for each problem, and
3.   Writing test cases and questions as well as setting scoring keys for each key feature.

Steps 1 and 2 are done first (often during one test committee meeting), followed later by Step-3 (during a subsequent test committee meeting), thus keeping separate the process of defining key features and preparing the test material per se.

4.1   PROBLEM SELECTION AND CONTEXT DEFINITION

This first step contains three sub-steps:

1.   Selecting the problems from the domain,
2.   Selecting the patient's age group, and
3.   Selecting the clinical situation and site of care.

Problem selection. To ensure content validity of a key features-based clinical decision-making examination, it is imperative to select a representative sample of problems from the entire spectrum of clinical problems for which the candidates are held accountable. For the MCCQE Part I, this spectrum (technically called the domain) is provided by the statements of presentation-based objectives that the MCC produces for its Qualifying Examination, the MCCQE objectives. Within these objectives are lists of clinical presentations and problems that a graduating student should be

competent in assessing. Each presentation or problem is expressed as a presenting complaint (e.g., abdominal pain) or a clinical problem (e.g., abdominal mass) and is accompanied by a list of pertinent diagnoses such as appendicitis, myocardial infarction, pneumothorax, migraine, cystitis, acute gastroenteritis, etc. Collectively the clinical presentations listed in the MCCQE objectives are those from which a sample is drawn for the key features-based CDM examination. A blueprint is used to determine which problems from the domain will be selected for inclusion on a given examination. Different strategies can be used for this blueprinting process. For example, one might select the problems in equal proportion from the major clinical disciplines (i.e., pediatrics, medicine, surgery, obstetrics-gynecology, psychiatry, and public health) or one might select based on the proportion of health services provided to different age groups in the population.

Patient age group. Most of the problems in the domain can occur at different times in a patient's life span, with varying manifestations and issues depending on the age of the patient. Five different age groups can occur:

1. Pregnancy, neonatal and infant (up to 1 year),
2. Pediatrics (1-11 years),
3. Adolescent (12-18 years),
4. Adult (19-64 years), and
5. Geriatrics (65 and older).

Clinical situations and site of care. The clinical problems on a CDM examination are described in the way in which the problem would be presented by a patient to a physician and can be classified in one or more of the following clinical situations:

- an undifferentiated complaint;

- a single typical problem;

- a multiple or multi-system problem;

- a life-threatening event; and

- preventive care and health promotion.

Only one, sometimes two, of these clinical situations are chosen for each problem to be assessed. The one(s) selected will influence the definition of the key features for that problem. At the level of a graduating medical student, it is advisable to select primarily typical presentations of clinical problems. When atypical presentations are used, almost all the candidates will fail to effectively resolve them and the problem loses its discrimination capability.

Finally, the site of care can also influence clinical actions and decisions because different resources are available in different venues. For example, in a case of acute chest pain, an immediate cardiology consultation and specialized investigations may not be available in an office setting compared to the emergency room.

Hence, the problem selection and context definition process begins with (i) the selection of a problem from the list of MCCQE objectives (e.g., abdominal pain, acute), followed by (ii) the selection of a patient's age group (e.g., adult), (iii) a clinical situation (e.g., life-threatening event), and (iv) the site of care (e.g., outpatient department). Only once these elements have been selected and defined can the development of key features proceed.

4.2   KEY FEATURES DEFINITION

The process of defining key features for a given problem can be approached in several ways. One straightforward approach is to select a clinical problem (e.g., shortness of breath on exertion), an age group (e.g., adult), a clinical situation (e.g., an undifferentiated complaint), and a site of care (e.g.,

office) and then address the question, "What are the critical steps in the resolution of this problem for patients of this age group presenting in this way?" Here is an example of a set of key features for a problem of 3rd trimester bleeding presenting in the emergency room.

Example of a set of Key Features for a Problem.

Given a pregnant woman experiencing third-trimester vaginal bleeding with no abdominal pain, the candidate will:

1. Consider placenta previa as a leading diagnosis;
2. Avoid performing a pelvic examination (because it may cause fatal bleeding);
3. Avoid discharging home from the emergency department; and
4. Order pelvic ultrasound (to confirm placenta location).

_____

Each key feature has three basic elements:

1. ***a set of conditions:*** "Given a pregnant woman experiencing third-trimester vaginal bleeding with no abdominal pain";
2. ***a task*** (i.e., a clinical action or decision): "Consider placenta previa as a leading diagnosis"; and
3. ***some qualifier:*** "leading diagnosis."

While every key feature must have the first two elements, the third element (qualifiers) is not always present, as illustrated in key features 2 to 4 in the example above. The elements in parentheses in the set of key features in the example above are not tasks per se but are included as part of the key feature to remind the test developers of the reason or rationale for the task.

This approach to defining key features can often lead to a narrow statement of a problem's key features, as it is specific to only one presentation of the problem. A more general approach to defining key features can be used by addressing the following question for a given problem and age group***, "What are the critical steps in the resolution of this problem for patients in this age group?"*** In answering this question, it is extremely useful to think of a variety of instances of patients in which the clinical problem can present itself (e.g., as an undifferentiated complaint, as a life threatening event, as a multiple problem, etc). This second approach to defining key features may be more productive in defining a broader set of key features, some of which are common to different clinical presentations of the problem. Furthermore, this approach can be further generalized by not specifying an age group and then including age as part of the "given clinical information" in the statement of the key features. When key features are more broadly defined in these ways, and a case is developed to portray this problem on the examination, it may turn out that not all key features defined are relevant to the case, and thus cannot be tested. Experience has shown that two to three key features can typically be tested by one case - that is, there are two to three critical steps in its resolution. Some cases have been found to possess as few as one key feature and others may have as many as five. From a psychometric perspective, 2 to 3 key features (or questions) per case optimizes reliability and efficient use of limited examination time (Norman et al, 2006).

The second more general approach to defining key features outlined above is preferable. In summary, it entails the following steps:

1. Select a clinical problem from the list of MCCQE objectives and an age group.
2. Pose the question, "What are the critical steps in the resolution of this problem for patients in this age group?" "Where's the challenge for a graduating medical student?"

3. Think of many instances or patients who present in different clinical situations (i.e., an undifferentiated complaint; a single typical problem; a multiple or multi-system problem; a life-threatening event; or preventive care and health promotion) with this problem.
4. List the key features for these patients, some that will be unique to specific patients, others that will be common across patients.
5. Select the 2 to 3 most critical key features.

The development of good key features can be challenging and the opportunity to present key features for review and refinement to colleagues or to other test committee members is a necessary step in the process, especially for authors inexperienced in defining key features. Novice authors often possess an initial reluctance to give up testing all aspects of the problem, from presenting complaint to follow-up, in contrast to selectively focusing on only the few critical decisions and actions and consequently giving credit only for those decisions or actions. Discussions of key features with colleagues should focus on such questions as, "Is the critical issue in this problem one of interpreting the data, or is it one of eliciting the data?" This distinction will result in asking very different questions on the examination. If interpretation is the issue, then clinical data are given and the examinee is asked questions aimed at interpretation (see *Example 5.1 Clinical Problem #1*). If data acquisition is the issue, then a complaint is given and the examinee is asked to identify the specific data that would be elicited from the patient (see *Example 5.2 Clinical Problem #2*). The clarity of a key feature is as important as the clarity of a research question in a research study or an instructional objective in a teaching session. If the key feature is clear, all other aspects of case development will readily unfold.

The same issues come up when asking for a diagnosis. Is the key feature focusing on (a) eliciting a probable diagnosis, that is, one diagnosis in a differential (as in an undifferentiated complaint or typical case presentation) or (b) eliciting THE working diagnosis or leading diagnosis at this time (as in a life-threatening situation)? The difference between the two key features is that in the first instance (a) the test question will ask for a differential diagnosis whereas in (b) the question will ask for one leading diagnosis at that point in time.

The process of defining key features can be frustrating if problems are defined too broadly. For example, it would be difficult to define key features for the general problem areas of "cardiovascular ischemic disorder" and "alcohol abuse," but not for problems of "stable and unstable angina" and "alcohol withdrawal." Alter-natively, selecting a single patient as the basis for defining key features provides too narrow a view of the problem and may inappropriately restrict the definition of key features.

Each key feature statement contains two or three basic elements:
- Initial clinical information (**conditions**) that casts the problem, patient's age, and clinical situation (i.e., "Given an adult patient presenting with … the candidate will"). The initial clinical information can consist of signs or symptoms, diagnoses, or laboratory test results (e.g., "Given a patient presenting with a swollen and/or painful leg …") depending on the nature of the key features tested.

- A clinical *task* (e.g., "…order a venogram…").The clinical task is a critical step, decision, or action, that is, the key feature, in the diagnosis or management of the problem (e.g., "… include DVT in a differential diagnosis" or "… elicit risk factors" or "… order a venogram").

- if necessary, a *qualifier*(s) may be added (e.g., "order an *initial* work up…, "order an *immediate* antibiotic," or "list the *most likely* diagnosis(es)"").

Key feature steps or actions need to be stated as specifically as possible (e.g., "prescribe morphine" or "order an arterial blood gas"). One should avoid non-specific steps or actions, such as "manage appropriately" or "order appropriate investigations." Similarly replace "exclude ongoing hypoxia," a statement of goal, by "order blood gases," that is, a specific clinical action. In addition, avoid

combining different actions or decisions into one key feature statement, such as "include DVT in a differential diagnosis and order a venogram". This statement should be separated into two distinct key features.

Sometimes a key feature can focus on actions to avoid; for example, "Avoid performing a pelvic examination (because it may cause fatal bleeding)" or "Avoid discharging home from the emergency department."

Finally keep in mind the level of experience or training of the candidates being assessed. The very same problem and conditions can yield different key features depending on whether one is assessing graduating students versus senior residents, or fellows.

4.3   CASE SCENARIO

After defining the problem's key features, a clinical case scenario must be developed to represent the problem and its key features on the examination. Case scenarios can vary in length depending on how much information is needed to get to the question testing the first key feature. For problems whose key features emphasize the diagnostic work-up of the patient, the scenario is often very brief. For problems in which the critical steps focus on laboratory or management issues, the scenario will typically be longer and contain clinical data derived from history taking, physical examination, and/or initial laboratory data. The usual elements common to any case scenario include the patient's age and gender, the setting (site) of the encounter, and the patient's presenting condition or reason for seeking care, followed by whatever clinical details are required prior to the presentation of the first question. When writing clinical scenarios, think of real patients you have seen in practice, rather than inventing them de novo. Using real patients or actual real-life situations will make the scenario more authentic (less contrived).

Use of lay language descriptions is also preferred over use of medical terminology or jargon ("medicalese"). In real life the information available to the candidates are lay terms (e.g., coughing up blood), not medicalese (i.e., hemoptysis). When clinical scenarios are presented using medicalese (such as "Homan's sign"), candidates, especially the weaker ones, are more likely to accurately identify the correct responses (Eva et al, 2010). The use of "medicalese" has the effect of "cueing" and decreases the discriminating ability of the clinical decision-making examination format. The use of lay language descriptions (such as "painful calf muscle") is more challenging for weaker candidates and increases the discrimination factor of the key feature being measured. The following is an example of a scenario using lay terminology.

Example of a clinical scenario.

"A 35-year-old mother of three sees you in the office at 1700h for watery diarrhea that started yesterday morning. She has had about 15 watery bowel movements in the past 24 hours, has felt sick to her stomach, but no vomiting. She works during the day as a cook in a restaurant but left work to come to your office. On her chart, your office nurse notes a resting blood pressure of 105/50 mm Hg lying down (a pulse of 110/minute), 90/40 mm Hg standing, and an oral temperature of 36.8 degrees C. On physical examination, you find she has a dry mouth and active bowel sounds. A urinalysis shows a normal microscopy, with a specific gravity of 1.030."

Also, in the initial scenario, it is important to put the candidates in a realistic situation and role. For example, instead of saying, "A 29-year-old woman with a red eye that is getting worse for two days…," write, ""You are asked to see a 29-year-old woman in the walk-in clinic. She tells you that her right eye has become more and more red over the past two days". When asking questions,

instead of asking, "Which diagnosis(es) should you consider?," write "What is (are) your diagnosis(es) at this time?". By going from "should" to "What is(are)", you are putting the candidates in their clinical role with a patient rather than asking them to come up with "the" correct answer as implied in the "should". Finally, be concise in writing the scenario, while at the same time providing all the necessary information leading up to the first question (KF); in other words, avoid useless information in the scenario. Unnecessary information takes time to read and is a poor use of testing time. However if the problem calls for extraneous yet pertinent information, make sure that that information is included.

Experience has shown that committee members who are familiar with the level of training of the candidates being tested can define very accurate and reliable key features. Clerkship directors from across Canada in a 1995 study corroborated 92% of the key features generated by CDM test committee members and generated de novo 94% of the committee's key features when given the problem, patient age, and clinical situation (Bordage, Brailovsky, Carretier, Page, 1995).

4.4   TEST QUESTIONS AND FORMATS

Generally, one question tests one key feature. Occasionally a question may be used to test more than one key feature (see *Example 5.1 Clinical Problem #1, Question #2*). The question stem consists of a direct formulation; for example, *"What is your leading diagnosis at this point in time?",* compared to incomplete sentences that would be completed by an examinee's responses (e.g., "Your leading diagnosis at this point in time is _____"). The questions derive directly from the key features and focus exclusively on clinical actions or decisions (as opposed to the reasoning behind the action or decision); for example, instead of saying "Consider contributing factors" write "Order thyroid function tests." A second sentence provides instructions regarding the number of permissible responses, for example, "List up to four". If the question is not the first in a series of questions pertaining to the clinical scenario, it may also contain additional information, for example, "The patient has now been in the hospital for two days, and his condition ... What treatment would you now recommend?".

Two response formats are recommended for clinical decision-making cases: short-answer "write-in" (WI) responses and "short-menu" (SM) responses. In the WI format, examinees supply their responses. In the SM format, examinees select their responses from prepared lists of options accompanying the question. The number of options in these lists varies depending on what the question is testing. In general there are typically 15-20 options, but the numbers can range from as few as 5 or 6 to as many as 30. A typical list of options for lab investigations would contain the list of lab tests usually found on a lab order sheet. To avoid cueing, the lists of options should be presented in ascending alphabetical order. The lists of options must, of course, contain all keyed correct responses plus incorrect responses and common misconceptions (i.e., distractors).   The number of options should be sufficient to reduce guessing effects. The lists may also contain correct options that are not scored as they are not integral to the key feature(s) being assessed. Such correct non-scored options serve as neutral distractors and must be taken into consideration when determining the total allowable number of responses. There is no steadfast rule about the number of options to include. Try to keep the number within the range of what would be plausible in real life, including plausible but incorrect options (e.g., example, include "blood glucose" in the list of options because a "blood glucose" test is a correct action but is NOT part of the key feature, and consequently will not be worth any points), especially options that would be attractive to weaker candidates. An excellent way to generate and refine option lists for SM questions is to pilot test the question as a WI question and then use the examinees' incorrect responses as a source of "distractor" options for the short menu.

When listing each option, include only one action (decision) per option. For example, instead of "Administer one dose of ampicillin immediately in the ER and discharge patient with a prescription for

oral ampicillin", split into two options: "Administer one dose of ampicillin immediately" and "Discharge patient with a prescription for oral ampicillin".

Almost always, the last option on the list will be "None," meaning that no action is needed at that particular time in the case. This "None" option is especially useful when testing over-investigation or over-treatment. By having the "None" option always present at the end of each list, the candidates are not overly cued when the option is present. Examples of "none" statements:

> Select (write) "none" (or "Do nothing") if no action [investigation, treatment] is needed at this point.
> Select (write) "none" (or "Do nothing") if you choose no action at this point.
> Select (write) "none" if none are warranted.

It is generally recommended to use the WI format for questions that test diagnoses, management, and prescriptions (see *Example 5.3 Clinical Problem #3*, Questions #1 and #2) while the SM format is generally used for history and physical exam findings or investigations (see *Example 5.3 Clinical Problem #3*, Questions #3 and #4). Questions testing for diagnoses and treatments are also most readily marked in the WI format, which is not the case for questions testing issues of history taking and physical examination because examinees can phrase responses in many equivalent ways. Questions that test issues of laboratory investigation are naturally presented in a SM "laboratory sheet" format (see *Example 5.3 Clinical Problem #3*, Question #4). Studies of the MCCQE Part I CDM examination data have shown that WI questions are more discriminating than SM questions, especially around the pass-fail cut-score, and therefore are more effective at identifying weaker candidates (Page, Boulais, Blackmore, Dauphinee, 2000).

While some problems will only have one key feature, the number of single key feature cases should not exceed 30% of the cases on an examination. The optimal number of key features per problem is 2 to 3 (Norman, Bordage, Page, Keane, 2006). Single-KF cases tend to be less reliable. On the other hand, asking 4 or more questions per case tends to be redundant and does not improve test score reliability.

The sample problems in Section 5 of these guidelines illustrate both WI and SM formats.

Having standard questions on the exam will help reduce biases that might come from different formulations of the same question.   Thus uniform formulation of questions is desirable. See examples below; statements in brackets represent alternatives.

Examples of test questions.

History & physical examination

- What element(s) of the history [or physical exam], if any, will you want to gather [ask about; elicit, focus on]?
- What focused questions will you ask [or look up in the chart] at this time?
- What features on history and physical exam, if any, will you look for at this time?
- What [pertinent] clinical information do you want at this time?
- As you review the chart and take her history, what are you particularly interested in?
- What else, if any, will you ask him [her] at this time?

Diagnosis

- What is your leading [working] diagnosis at this point? List only one.
- What diagnosis(es) are you considering [thinking about] at this time? List up to three.
- The chest radiograph is now available (show film of …). What is your diagnosis?

- What is the [most likely] [specific] cause of this patient's condition?

Investigation

- What investigation(s) [test(s)], if any, will you order at this point?
- What test(s), if any, will you order next?
- What further [additional] investigation(s), if any, will you order?

Management

- How will you manage [treat] this patient at this point?
- What are your orders at this time? Be specific.
- What action(s) [step(s)], if any, will you take?
- What, if anything, will you do at this time [next]?
- What order(s), if any, will you write for the next few hours?
- What is the next [essential] step(es) in your management?
- What is your immediate management of this patient at this time? Be specific.
- What steps will you take in your immediate assessment and management of this patient?
- After initial management of this patient's condition, what other measures, if any, will you order?
- A stat bedside blood sugar is 32 mg/dL. You will now ask the nurse to:…
- What management option(s) will you offer [recommend to] your patient at that time?
- What treatment [measures] [additional measures] will you recommend? Be specific.
- What advice [information] will you give the patient?
- What will you recommend as the next step(s) in the management of this patient?

Response limits. Every question is followed by one of the following statements that sets the limit of allowable responses:

- List (select) only one.
- List (select) up to x.
- List (select) as many as are appropriate.

Candidates are warned about these different types of limits in the general instructions to the exam (see Section 4.8).

- List (select) only one.

> This type of limit is suitable for requesting a single definitive answer, such as a leading diagnosis or a most important management step.

- List (select) up to x.

> This type of limit is suitable for questions in which one or more answers are sought and for which the number of opportunities for examinees to provide these answers should be capped (i.e., forcing the issue). The "x" (number of allowable responses) takes into account the number of correct keyed responses sought, together with other factors such as responses that would be reasonable but not in the keyed responses as well as incorrect responses that might be appealing to weaker candidates.

- List (select) as many as are appropriate.

This type of limit is used in situations where it is useful to determine how many actions an examinee might take, as in the case of limiting the investigation or treatment (i.e., not over investigating or treating). The maximum number of responses is specified in the scoring key, but the examinees are blinded to this limit.

Finally, it is very useful to include a summary table illustrating the relationship between test questions and key features. At a glance the test developer can see the layout of the case and its questions and key features. For example in the table below, the case contains 2 questions that test 3 key features.

	KF-1	KF-2	KF-3	KF-4
Q1	X			
Q2		X	X	
Q3				
Q4				

It is usually possible to test all key features using one case scenario, but again there are exceptions. Issues of excessive cueing across questions make it necessary, at times, to select a second case and develop a second case scenario to test the remaining key features. The second case would be separated from the first during the delivery of the examination, and presented in such a way that it would not be readily recognized as related to the first.

4.5   SCORING KEYS

The scoring key consists of the list of the correct responses to a question and a system of assigning numerical scores to the keyed responses. The correct scored responses should directly reflect the key feature(s) being assessed. All the correct answers appearing in the scoring key should be part of the key feature(s), no more no less. Some scoring keys contain only a single response, others contain several responses and, in addition, the responses may be clustered on the basis of logical considerations regarding the correct clinical actions to be taken. See examples below.

Examples of Scoring Keys

Key Feature 1.   Given an adult patient presenting with a painful, swollen calf, the candidate will include deep venous thrombosis in the differential diagnosis.

Score	Criteria	Synonyms
1.0	Deep venous thrombosis	Deep vein thrombosis, DVT
0	Listing more than two responses	

Key Feature 3. Given an adult brought to the emergency room with multiple seizures and without having regained consciousness, the candidate will elicit a history to identify possible causes of the seizures (including alcohol use, medications, illicit drugs, and diabetes history).

Score	Criteria	
1	0.25	#2. Alcohol history
	0.25	#5. Cocaine abuse or #15. Heroin abuse
	0.25	#7. Diabetes history
	0.25	#18. Medication history
0	#33.	Not appropriate to call at this point in time
	or	
	Selecting more than six items	

Scoring keys for each key feature define the responses needed to produce a score of "1" for the key feature. Scoring keys assign a weight to each correct response – such as the same weight of 0.25 to each of four correct responses. Alternately, some responses could be considered more important than others and assigned larger weights. As a general rule, use of equal weighing is preferred (see Examples **5.3 Clinical Problem #3**, Questions #3 and #4), since research has shown that differential weighting (see Example **5.3 Clinical Problem #3**, Question #2) does not improve score reliability and is therefore an inefficient use of test development time determining the weighted scores. Some scoring keys also list responses that result in a score of "0" no matter what other responses an examinee makes (see Examples **5.1 Clinical Problem #1**, Question #2 and **5.4 Clinical Problem #4**, Question #2). These so-called "killer" responses are used for selecting or listing too many responses or because they represent actions that are dangerous or harmful to the patient (e.g., unnecessary invasive investigation or harmful treatment). Consequently, not all dangerous or harmful actions will result in a score of "0."   For example, if the candidate (e.g., a graduating medical student) orders an unnecessary angiography on the examination, in real life, the radiologist or the cardiologist would review and cancel the order before it would be executed. The inclusion of too many "killer" options will have a negative effect on the discrimination of the question. On the other hand, some harmful actions or decisions can be executed without any safety net present, like discharging a patient from the emergency room in the middle of the night, or prescribing a contra-indicated medication, or executing potentially harmful maneuvers (like doing a vaginal examination for a women with 3rd trimester bleeding), and should result in a score of 0 for the key feature.

Whatever weights are assigned to keyed correct responses, they should sum to "1" so that each question within a problem is given a maximum score of "1" and a minimum score of "0". When an examination question tests more than one key feature, each key feature is given a score of 1. Thus key features, not questions, are assigned points. See example below.

Example of a question testing two Key Features.

Question 2. After management of the patient's acute condition, what additional measures, if any, would you recommend? Select up to four or select "None" if none is appropriate.

Key Feature 2. Given an adult patient with gastroenteritis and dehydration, the candidate will collect clinical specimens

Key Feature 3. Given an adult patient with gastroenteritis and dehydration, the candidate will recommend exclusion of food handler from the workplace until asymptomatic

Scoring Keys
KF-2

Score	Criteria
1.0	#8. Stool cultures
0	#11. None

Selecting more than four items

KF-3

Score	Criteria
1.0	#10. Temporary absence from work
0	#5. Give immune serum globulin to close family contacts
	#11. None
	Selecting more than four items

Key feature scores within problems are then averaged so that each clinical case on the examination has a maximum score of "1". Examination scores are then the sum of case scores divided by the total number of cases on the examination. This summing and averaging procedure ensures that key features have equal weighting within a case and that cases have equal weighting across the entire examination.

The passing score for a test is based on a content, criterion-based standard setting procedure, using a modified Angoff procedure that takes about 30 minutes per problem (see Norcini (2003) for a review of standard setting procedures).

4.6   DOCUMENTATION

For each problem and corresponding set of key features, include one or more references to support the key features. For example, for the Seizures problem (Epilepsy - 92) (see **Example 5.3 Clinical Problem #3**), the following reference was used, especially regarding the use of thiamine:

Sirven, JI & Waterhouse, E. Management of status epilepticus. Am Fam Phys 2003; 68:469-76.

4.7   TESTING TIME AND TEST SCORE RELIABILITY

On average it takes candidates 3 minutes to respond to each question and 6.5 minutes per case. For a 3.5-hour exam, 32 cases can be tested (i.e., about 71 key features with 2-3 key features per case) (Bordage, Carretier, Bertrand & Page, 1995).

On average, the reliability of test scores for a 3.5-hour exam (containing 32 cases) is about 0.70 (Cronbach's alpha). This represents a huge increase in reliability compared to Patient Management Problems where about 10 cases were tested for the same amount of time, each case testing everything from chief complaint to history & physical examination, investigation, management, and follow-up. Using the Spearman-Brown Prophecy Formula, it would take about 45 to 50 cases (a one-day exam) to reach a 0.80 level of reliability. It is important to note that test score reliability is calculated using the case as the unit of measurement (e.g., 32 cases) and not the individual key features (e.g., 71 key features).

4.8   INSTRUCTIONS TO CANDIDATES

Given the peculiarities of the key features approach, especially regarding scoring and the fact that once a case is completed the candidate can't go back to change responses, candidates are provided with clear instructions and practice cases on how to take this type of examination. For the complete set of instructions, see: www.mcc.ca/english/examinations/qualifying_e1.html

On the left-hand side of every examination page the following information and links are available to the candidates:

1.   link to the French version for English candidates (or the opposite if the candidate is taking the examination in French),
2.   a clock showing the amount of time remaining for the examination,
3.   link to normal values (candidates are expected to know the normal values for common tests but all the normal values are available within this link),
4.   link to a calculator, and
5.   a submit button (after which the candidate cannot go back to review or change any response).

It is essential for candidates taking this type of examination to read the instructions carefully and practice before showing up for the examination. Overall candidates do well in navigating this type of examination.

4.9   DISSEMINATION

The Key Features approach to assessment has been widely discussed (Bowles, 1995; Norman, Swanson, Case, 1996) and provides the following psychometric advantages:

- **Better reliability** because of adequate sampling of the domain, that is, 30-35 short problems for a 3.5-hour examination;
- **Better content validity** because of more precise assessment of the important clinical decisions for a representative sample of the domain, especially in assessing bottom-up reasoning, that is, assessing knowledge application, what clinicians do;
- Use of varied **response formats** to best assess particular decision skills, and focused and more precise **scoring**;
- Justifiable **pass-fail decisions**; and
- **Best predictor** of future complaints (Tamblyn et al, 2007).

Beyond all Canadian medical schools, the key features approach is used worldwide to assess clinical decision making. For example, the following institution and testing agencies use the key features approach:

Collège des Médecins du Québec - Structured Oral Interviews (Jacques et al, 1995; Miller et al, 1997)

College of Physicians & Surgeons of Pakistan (Ali, Bordage, 1995)

American College of Physicians (MKSAP recertification)

American College of Colon & Rectal Surgeons - CARSEP recertification (Trudel, Bordage & Downing, 2008: 9 cases; Cronbach's ☐ =.95 overall .93 Colo-rectal surgeons)

Royal Australian College General Practitioners (Farmer & Page, 2005; Farmer, Hinchy, 2005)

Swiss National Examination Board

McMaster clerkships (Hatala & Norman, 2002: 15 cases; Cronbach's ☐=.49)

German medical schools (Fischer, Kopp, Holzer, Ruderich, Junger, 2005).

The approach has also been used for curriculum and instruction purposes (Doucet, Purdy, Kaufman, Langille, 1998; Sturmberg, Crowe, Hughes, 2003).

V.   EXAMPLES OF KEY FEATURES PROBLEMS

5.1   CLINICAL PROBLEM #1

Problem   1

Author   Dr. X

Clinical Problem   Third Trimester Bleeding

MCC Objective   Vaginal Bleeding (112); Pregnancy - Obstetrical Complications (80-3)

Life Span Period

- ‿   pregnancy/neonatal/infant (up to 1 year)
- ‿   pediatric (1-11 years)
- ‿   adolescence (12-18 years)
- ✓   adult (19-64 years)
- ‿   geriatric (over 65 years)

Clinical Situation

- ‿   undifferentiated complaint
- ✓   single typical problem
- ‿   multiple or multisystem problem
- ✓   life-threatening event
- ‿   preventive care and health promotion

Location/Setting   Emergency Department

Patient's Age   20-30 years

Patient's Gender   Female

Key Features

Given a pregnant woman experiencing third-trimester vaginal bleeding with no abdominal pain, the candidate will:

1.   Consider placenta previa as a leading diagnosis;
2.   Avoid performing a pelvic examination (because it may cause fatal bleeding);
3.   Avoid discharging home from the emergency department; and
4.   Order pelvic ultrasound (to confirm placenta location).

	KF-1	KF-2	KF-3	KF-4
Q1	X			
Q2		X	X	X

Case Scenario

At 20:00h, the nurse in the emergency department asks you to see a 24-year-old woman, who has 2 children and is now 31 weeks pregnant, presents with bright red bleeding coming from her vagina that began about two hours ago. The three sanitary napkins that she used were completely soaked.

Her pregnancy has been normal, as were the two previous ones. She has not had any pain in her belly and she says that the fetus is moving as usual.

Her BP is 110/70 mm Hg and her pulse is 92/min. Your examination of the abdomen reveals a uterine height of 31 cm with a soft, non-tender uterus. The fetus is in a breech position and has a heart rate of 150/min. No bleeding has occurred since she arrived 25 minutes ago.

Question 1   Format WI

What is your leading diagnosis at this time? List only one or write "normal" if you judge this situation to be within normal limits.

1. _____

Key Feature

1. Given a pregnant woman experiencing third trimester vaginal bleeding with no abdominal pain, the candidate will consider placenta previa as a leading diagnosis.

Scoring Key

Score	Criteria	Synonyms
1.0	Placenta previa	Marginal placenta, low placenta, low insertion
0	Listing more than one response	

Question 2   Format SM

What steps will you take in your immediate assessment and management of this patient? Select as many as are appropriate.

1. Artificial rupture of membranes
2. Cervical swab for chlamydia
3. Complete blood count
4. Computed tomographic scan of abdomen & pelvis
5. Cross and match for 2 units of blood
6. Discharge home with instructions to come back if bleeding starts again
7. Fetal scalp monitor to assess status of fetus
8. Human immunodeficiency virus serology
9. International normalized ratio (INR)
10. Large bore intravenous line
11. Partial thromboplastin time (PTT)
12. Pelvic ultrasound
13. Pelvic examination
14. Syphilis serology
15. Ultrasound using vaginal probe
16. Vaginal swab for Group B streptococcus

Key Feature 2.

Given a pregnant woman with third trimester vaginal bleeding with no abdominal pain, the candidate will avoid performing a pelvic examination because it may cause fatal bleeding.

Score	Criteria
1.0	Did NOT select #13. Pelvic examination
0	#1. Artificial rupture of membranes #2. Cervical swab for chlamydia #7. Fetal scalp monitor to assess status of fetus #13. Pelvic examination or Selecting more than six items

Key Feature 3.

Given a pregnant woman with third trimester vaginal bleeding with no abdominal pain, the candidate will avoid discharging home from the emergency department.

Score	Criteria
1.0	Did NOT select #6. Discharge home with instructions to come back if bleeding starts again
0	#6. Discharge home with instructions to come back if bleeding starts again or Selecting more than six items

Key Feature 4

Given a pregnant woman with third trimester vaginal bleeding with no abdominal pain, the candidate will order pelvic ultrasound (to confirm placenta location).

Score	Criteria
1.0	#12. Pelvic ultrasound
0	#15. Ultrasound using vaginal probe or Selecting more than six items

5.2   CLINICAL PROBLEM #2

Problem   2

Author   Dr. Y

Clinical Problem   Painful Swollen Leg

MCC Objective   Localized Edema (29-2)

Life Span Period
- ＿ pregnancy/neonatal/infant (up to 1 year)
- ＿ pediatric (1-11 years)
- ＿ adolescence (12-18 years)
- ✓ adult (19-64 years)
- ＿ geriatric (over 65 years)

Clinical Situation
- ＿ undifferentiated complaint
- ✓ single typical problem
- ＿ multiple problem or multisystem problem
- ＿ life-threatening event
- ＿ preventive care and health promotion

Location/Setting   Outpatient Clinic

Patient's Age   Older adult (55-65 years)

Patient's Gender   N/A

Key Features

Given an adult patient presenting with of a painful, swollen calf, the candidate will:

1. Include deep venous thrombosis in the differential diagnosis
2. Elicit a history of risk factors for deep venous thrombosis: cigarette smoking, previous knee problems, previous neoplasia, recent intercontinental air travel, recent surgery, and recent work environment

	KF-1	KF-2	KF-3	KF-4
Q1	X			
Q2		X		

Case Scenario

A 56-year-old man sees you in the outpatient clinic because of pain in his left lower leg that began two days ago and has been getting worse since. His ankle is also swollen. He has never had a similar problem. His other leg is fine.

Question 1   Format WI

What diagnosis(es) are you considering at this time? You may list up to two.

1. _____

2. _____

Key Feature 1.

Given an adult patient presenting with a painful, swollen calf, the candidate will include deep venous thrombosis in the differential diagnosis.

Scoring Key

Score	Criteria	Synonyms
1.0	Deep venous thrombosis	Deep vein thrombosis, DVT
0	Listing more than two responses	

Question 2   Format SM

With respect to your diagnosis(es), what elements of his history would you particularly want to elicit? You may select up to seven. Select option #27 if no other information is needed at this time.

1. Alcohol intake
2. Allergies
3. Angina pectoris
4. Anti-inflammatory medication
5. Cigarette smoking
6. Colour of stools
7. Cough
8. Headache
9. Hematemesis
10. Impotence
11. Intermittent claudication
12. Low back pain
13. Nocturia
14. Palpitations
15. Paresthesias
16. Paroxysmal nocturnal dyspnea
17. Polydipsia
18. Previous back problems
19. Previous knee problems
20. Previous neoplasia

21.   Previous urinary tract infection
22.   Recent dental procedure
23.   Recent intercontinental air travel
24.   Recent sore throat
25.   Recent surgery
26.   Recent work environment
27.   No other information is needed

Key Feature 2.

Given an adult patient presenting with a painful, swollen calf, the candidate will elicit risk factors for deep venous thrombosis through the patient's history

Scoring Key

Score	Criteria
1.0	Must select at least five of the following six items:
	#5. Cigarette smoking
	#19. Previous knee problems
	#20. Previous neoplasia
	#23. Recent intercontinental air travel
	#25. Recent surgery
	#26. Recent work environment
0	Not selecting five of the above six items
	or
	Selecting #27. No other information is needed
	or
	Selecting more than seven items

5.3   CLINICAL PROBLEM #3

Problem   3

Author   Nendaz, M.R., Bordage, G., University of Illinois in Chicago.

Clinical Problem   Seizures

MCC Objective   Seizures (Epilepsy) (92)

Life Span Period

_   pregnancy/neonatal/infant (up to 1 year)

_   pediatric (1-11 years)

_   adolescence (12-18 years)

✓ adult (19-64 years)

_ geriatric (over 65 years)

Clinical Situation

_ undifferentiated complaint

✓ single typical problem

_ multiple problem or multisystem problem

✓ life-threatening event

_ preventive care and health promotion

Location/Setting   Emergency Department

Patient's Age   N/A

Patient's Gender   N/A

Key Features

Given an adult brought to the emergency room with multiple seizures and without having regained consciousness, the candidate will:

1. Generate a provisional diagnosis of status epilepticus;
2. Begin initial therapy including administration of intravenous vitamin B complex (thiamine), hypertonic glucose bolus (50% dextrose), and lorazepam or diazepam or clonazepam or phenobarbital AND phenytoin;
3. Elicit a history to identify possible causes of the seizures (including alcohol use, medications, illicit drugs, and diabetes history); and
4. Order immediate investigations to identify potentially treatable causes of the seizures, including electrolytes, glucose, calcium, arterial blood gases, drug screen, and CT or MRI of brain.

	KF-1	KF-2	KF-3	KF-4
Q1	X			
Q2		X		
Q3			X	
Q4				X

Case Scenario

A 36-year-old man is brought to the emergency room by ambulance because he fell onto a sidewalk unconscious while waiting for the bus. A witness immediately called an ambulance and reported to

the ambulance crew that before falling to the ground, he seemed confused, agitated, and was arguing with some invisible person. After falling, he began to twitch for a short while, his face became blue, and then he began to have jerky movements all over his body for about a minute. He did not recover consciousness after the episode. During the 10-minute ambulance trip, he presented two other similar episodes, without recovering consciousness, and a third episode that you witnessed on arrival in the emergency room.

His temperature is 37.8 degrees C. He looks neglected and is unconscious. No relatives or friends accompanied the patient.

Question 1   **Format WI**

What is (are) your leading working diagnosis(es) at this point in time? You may list up to two.

1.   _____

2.   _____

Key Feature   1.   Given an adult brought to the emergency room with multiple seizures and without having regained consciousness, the candidate will generate a provisional diagnosis of status epilepticus

Scoring Key

Score	Criteria
1.0	Status epilepticus (Note: both elements required)
0	Listing more than two responses

Question 2   **Format WI**

What is your immediate management at this point in time (be specific)? List as many as are appropriate.

1.   _____

2.   _____

3.   _____

4.   _____

5.   _____

6.   _____

7.   _____

8.   _____

9.   _____

10.   _____

Key Features 2.

Given an adult brought to the emergency room with multiple seizures and without having regained consciousness, the candidate will begin initial therapy including administration of intravenous vitamin B complex (thiamine), hypertonic glucose bolus (50% dextrose), and lorazepam or diazepam or clonazepam or phenobarbital AND phenytoin

Scoring Key

Score	Criteria	Synonyms
0.25	IV Thiamine (must mention IV)	IV Vitamin B complex
0.25	IV Hypertonic glucose (must mention IV)	50% dextrose IV
0.5	IV Lorazepam (must mention IV)	IV Diazepam, IV clonazepam, IV Ativan, IV Valium, or IV phenobarbital
	AND	
	IV Phenytoin (must mention IV)	IV Dilantin
0	Writing more than 7.	

Question 3   Format SM

Ten minutes after arrival, the patient is still unconscious. The nurse found a telephone number in his wallet. What questions will you ask the person answering the phone – assuming that he/she knows the patient? Select up to six questions or select option 33 if you think that it is not appropriate to call at this point in time.

1. Abdominal pain
2. Alcohol history
3. Back pain history
4. Cancer history
5. Cocaine abuse
6. Coronary bypass history
7. Diabetes history
8. Diarrhea
9. Dizziness
10. Drug allergy
11. Family history
12. Food allergy
13. Headache
14. Hearing disability
15. Heroin abuse

16.   Joint pain
17.   Lung infection
18.   Medication history
19.   Muscular disease
20.   Nausea
21.   Palpitation history
22.   Pet in household
23.   Previous similar problem
24.   Profession
25.   Sexual history
26.   Smoking history
27.   Social integration difficulties
28.   Surgery
29.   Travel history
30.   Viral infection
31.   Visual impairment
32.   Vomiting
33.   Not appropriate to call at this point in time

Key Feature 3.

Given an adult brought to the emergency room with multiple seizures and without having regained consciousness, the candidate will elicit a history to identify possible causes of the seizures (including alcohol use, medications, illicit drugs, and diabetes history)

Scoring Key

Score		Criteria
0.25		#2. Alcohol history
0.25		#5. Cocaine abuse
	OR	#15. Heroin abuse
0.25		#7. Diabetes history
0.25		#18. Medication history
0		#33. Not appropriate to call at this point in time
		Selecting more than six items

Question 4   Format SM

It has now been 15 minutes since the patient's arrival. What investigations would you order at this point? Select as many as are appropriate or select option 34 if you think that investigations are not needed at this point in time.

1.   Alanine aminotransferase (ALT)
2.   Alcohol level
3.   Aldolase, serum
4.   Alkaline phosphatase, serum

5. Amylase, serum
6. Arterial blood gases (ABG)
7. Aspartate aminotransferase (AST)
8. Brain computerized tomographic (CT)-scan
9. Brain magnetic resonance imaging (MRI)
10. Calcium, serum
11. Carotid ultrasound-doppler
12. Cerebral angiography
13. Cerebrospinal fluid examination
14. Complete blood count (CBC)
15. C-reactive protein
16. Creatine phosphokinase, serum
17. Creatinine, serum
18. Drug screening, serum
19. Drug screening, urine
20. Echovirus, serology
21. Electroencephalographic (EEG) recording
22. Electrolytes (Na, K, Cl)
23. gamma-Glutamyl transferase (GGT)
24. Glucose, serum
25. Lactate dehydrogenase, serum (LDH)
26. Lyme disease, serology
27. Protein electrophoresis, plasma
28. Syphilis, serology
29. T4, free
30. Temporal artery biopsy
31. Thyroid stimulating hormone (TSH)
32. Total protein, plasma
33. Urea, serum
34. No investigations needed at this point in time

Key Feature 4.

Given an adult brought to the emergency room with multiple seizures and without having regained consciousness, the candidate will order immediate investigations to identify potentially treatable causes of the seizures (including electrolytes, glucose, calcium, arterial blood gases, drug screen, and CT or MRI of brain)

Scoring Key

Score	Criteria
0.17	#6. Arterial blood gases (ABG)
0.17	#8. Brain computerized tomographic (CT)-scan
	or
	#9. Brain magnetic resonance imaging (MRI)
0.17	#10. Calcium, serum

0.17	#18.	Drug screening, serum
	or	
	#19.	Drug screening, urine
0.17	#22.	Electrolytes (Na, K, Cl)
0.17	#24.	Glucose, serum
0	#34.	No investigations needed at this point in time
		Selecting more than nine items

5.4 CLINICAL PROBLEM #4

Problem   4

Author   Dr. A

Clinical Problem   Diarrhea

MCC Objective   Acute Diarrhea (22-1)

Life Span Period
- _ pregnancy/neonatal/infant (up to 1 year)
- _ pediatric (1-11 years)
- _ adolescence (12-18 years)
- ✓ adult (19-64 years)
- _ geriatric (over 65 years)

Clinical Situation
- _ undifferentiated complaint
- ✓ single typical problem
- _ multiple problem or multisystem problem
- _ life-threatening event
- ✓ preventive care and health promotion

Location/Setting   Physician's Office

Patient's Age   N/A

Patient's Gender   N/A

Key Features
Given an adult patient with gastroenteritis and dehydration, the candidate will:

1. Initiate oral rehydration;
2. Collect clinical specimens; and
3. Recommend exclusion of food handler from the workplace until asymptomatic.

	KF-1	KF-2	KF-3	KF-4
Q1	X			
Q2		X	X	

Case Scenario

A 35-year-old mother of three sees you in the office at 1700h for watery diarrhea that started yesterday morning. She has had about 15 watery bowel movements in the past 24 hours, has felt sick to her stomach, but no vomiting. She works during the day as a cook in a restaurant but left work to come to your office. On her chart, your office nurse notes a resting blood pressure of 105/50 mm Hg lying down (a pulse of 110/minute), 90/40 mm Hg standing, and an oral temperature of 36.8 degrees C. On physical examination, you find she has a dry mouth and active bowel sounds. A urinalysis shows a normal microscopy, with a specific gravity of 1.030.

Question 1   Format SM

How will you manage this patient at this point in time? Select up to three.

1. Antidiarrheal medication
2. Antiemetic medication
3. Intravenous antibiotics
4. Intravenous fluids
5. Nasogastric tube and suction
6. Nothing by mouth
7. Oral antibiotics
8. Oral fluids
9. Rectal tube
10. Send home with close follow up
11. Surgical consultation
12. Transfer to hospital

Key Feature   1.

Given an adult patient with gastroenteritis and dehydration, the candidate will initiate oral rehydration

Scoring Key

Score	Criteria
1.0	#8. Oral fluids
0	Selecting more than three items

Question 2   Format SM

After management of the patient's acute condition, what additional measures, if any, would you recommend? Select up to four or select "None" if none is appropriate.

1. Avoid dairy products
2. Colonoscopy
3. Frequent handwashing
4. Gastroenterology consultation
5. Give immune serum globulin to close family contacts
6. Infectious disease consultation
7. Notify Public Health Authority
8. Stool cultures
9. Strict isolation of patient
10. Temporary absence from work
11. None

Key Feature 2.

Given an adult patient with gastroenteritis and dehydration, the candidate will collect clinical specimens

Scoring Key

Score	Criteria
1.0	#8. Stool cultures
0	#11. None Selecting more than four items

Key Feature 3.
Given an adult patient with gastroenteritis and dehydration, the candidate will recommend exclusion of food handler from the workplace until asymptomatic

Scoring Key

Score	Criteria
1.0	#10. Temporary absence from work
0	#5. Give immune serum globulin to close family contacts #11. None Selecting more than four items

VI.   CHECKLIST FOR THE DEVELOPMENT OF KEY FEATURES TEST CASES

Guidelines for Selecting Clinical Problem, Defining Key Features, and Preparing Test Materials (test scenario, test questions and scoring keys)
- Select the clinical problem or presenting complaint from the MCCQE list of objectives.

- Select gender and age range of the patient; only specify a precise age (e.g., 27) if the problem specifically calls for a precise age.

- Select a clinical situation(s) best suited to the chosen clinical problem:

    1. Undifferentiated problem or complaint
    2. Single typical or atypical problem
    3. Multiple problem or multisystem involvement
    4. Life-threatening situation
    5. Preventive care and health promotion.

- Select a site of care (e.g., office, hospital ward, emergency department)

Guidelines for Developing Key Features

- Determine the critical or essential clinical decision-making steps for the resolution of the chosen clinical problem; answer the question, ***"What are the critical steps or elements in the resolution of this clinical problem?" "Where is the challenge?"***

- Think of various patient presentations pertaining to the chosen clinical problem.

- Decide if the critical steps in resolving the problem require eliciting data, interpreting data, or managing the condition.

- Write specific key features pertaining to the chosen clinical problem.

- The key feature statement should contain the following parts:

    1. The initial clinical information (i.e., "Given an adult patient presenting with …, the examination   candidate will");
    2. The clinical task (e.g.., "…Order investigations including creatinine …"); and
    3. If necessary, a qualifier(s) may be added (e.g., "Order immediate creatinine")

- Avoid trying to assess all clinical decision making steps in the resolution of a clinical problem. Focus only on the more challenging or discriminating essential steps, the difficult steps in practice where clinicians tend to go wrong.

- Discuss and review key features with colleagues and fellow committee members before developing a clinical scenario(s) for the key feature(s).

Guidelines for Developing Clinical Case Scenarios

- Write short clinical scenarios if the key feature(s) is(are) primarily eliciting data; write longer clinical scenarios if key feature(s) is(are) requiring data interpretation or management.

- Use "real patient data" as much as possible; using factitious data can risk introducing inadvertent inconsistencies (i.e., incompatible laboratory values).

- Avoid the use of "medicalese" and technical jargon when writing the clinical scenario; use the words that the patient would actually say or describe the physical findings rather than giving the medical term (e.g., "finger nails and nail beds are normal" instead of "no clubbing.").

- Use of photos (i.e., skin rashes, radiographs, etc) is encouraged in the clinical decision making examination format.

- Avoid the use of humor as this can potentially distract examinees.

Guidelines for Writing Questions

- Select the appropriate question format (write-in [WI] format preferred for differential diagnosis and treatment questions; short-menu [SM] format preferred for eliciting history or laboratory investigations).

- Decide on the appropriate instructions to accompany the question (i.e., "Select up to x", "Select only one," or "Select as many as are appropriate")

- In the SM format, include choices that may contain correct non-scored responses (i.e., not a key feature element), incorrect distractors, and dangerous options that would automatically give a "0" score if chosen.

- Distractors should be plausible; distractors should be developed with the acceptably competent medical graduate in mind; distractors should include content which is reflective of common misconceptions or errors which are not acceptable.

- Acceptable distractors:

  - are usually homogeneous in content (e.g., all are investigations or all are therapies)

  - may include correct answers that are not integral to the key feature(s) (i.e., neutral non-scoring distractors)

  - do not contain any hints to the correct answer

  - would seem plausible and attractive to the minimally competent medical graduate or one selecting an inappropriate course of action

  - are similar to the correct answer(s) in construction and length

- In the SM format, the list of choices should appear in alphanumeric ascending order, ending most of the time with a "None" option.

- In the WI format, the total number of allowable answers should be limited to the expectation of how many reasonable opportunities a medical graduate should have to include the correct response(s) (for example, if the most likely diagnosis should be included within the top 3 options, then the total number of allowable write-in answers should be no more than 3).

- Avoid questions that can be answered without reference to the clinical scenario; such questions are likely not measuring clinical decision making.

Guidelines for Preparing the Scoring Key

- All keyed correct scored responses must directly match the key feature(s).

- Use equal weighting if there is more than one correct response. Avoid use of differential weighting because it does not improve the reliability of the scores.

- Each key feature receives a total score of "1"; hence, if more than one correct response within a key feature, weighted scores should add up to "1".

- If more than one key feature is being assessed by a question, the scoring key should give a score of "1" for each key feature that is scored in the answer key.

- Use of incorrect dangerous responses ("killers") that automatically receive a score of "0" regardless of other responses for that particular question should be used judiciously, as would be plausible in the given clinical situation (with no safety net) and selected by weaker candidates.

- Include acceptable synonyms (equivalences) for keyed correct responses in WI questions, to facilitate scoring of such questions.

VII. SAMPLE KEY FEATURES PROBLEM TEMPLATE

Key features problem specifications are broken down into 15 sub-sections. Each key feature should link to the MCCQE objective that is being addressed by the problem (e.g., "Consider placenta previa as a leading diagnosis" for the problem of third trimester bleeding is covered in MCCQE objectives Vaginal Bleeding #112 and Pregnancy - Obstetrical Complications #80-3). This provides information to help in the blueprinting of the examination and in determining what clinical problems may be lacking in the item bank.

The following list provides the outline for designing key features problems. Please refer to section 5 (*Examples of Key Features Problems*) for more details and examples for constructing key features problems. Committee members should submit key features problem materials using this template.

1.   *Problem Number:*

2.   *Author(s):*

3.   *Title of Clinical Problem:*

4.   *MCC Objective(s) from which Clinical Problem is derived:*

5.   *Life Span Period (check one):*
    _   pregnancy/neonatal/infant (up to 1 year)
    _   pediatric (1-11 years)
    _   adolescence (12-18 years)
    _   adult (19-64 years)
    _   geriatric (over 65 years)

6.   *Clinical Situation (check one or two):*
    _   undifferentiated complaint
    _   single typical problem
    _   multiple or multisystem problem
    _   life-threatening event
    _   preventive care and health promotion

7.   *Location, Site of Care:*

8.   *Patient's Age* (specify only if essential to the problem; otherwise say N/A)

9.   *Patient's Gender* (specify only if essential to the problem; otherwise say N/A)

10.   *Key Features* (each key feature should be linked to the MCCQE objective(s) identified in above):

11.   *Dates of Revisions* (list all dates, including initial date of problem development and any subsequent revision dates by month/year):

12.　　**Documentation** (to support key features):

13.　　**Case Scenario** (clinical case stem required before asking the first question)

14.　　**First and Subsequent Questions** (WI or SM format)

15.　　**Scoring Key following each question with accompanying key feature(s) being assessed**

VIII.　REFERENCES

- Ali SK, Bordage G. Validity of Key Features for a Family Medicine Pilot Exam at the College of Physicians and Surgeons Pakistan. J Coll Phys Surg Pakistan 1995, 5(6):256-60 .

- Bordage, G., Page, G.　An Alternative to PMPs: The "Key Features" Concept. Further Developments in Assessing Clinical Competence, 2nd Ottawa Conference, 1987, 59-75.

- Bordage G, Brailovsky C, Carretier H, Page G. Content Validation of Key Features on a National Examination of Clinical Decision-making Skills. Ac Med 1995; 70:276-81.

- Bordage G, Carretier H, Bertrand R, Page G. Comparing Times and Performances of French- and English-speaking Candidates Taking a National Examination of Clinical Decision-making Skills. Acad Med 1995; 70:359-365.

- Bowles, T.L.　A Worthy Search:　The Development of the Key-Features Concept.　Acad Med 1995;70:89-90.

- Doucet MD, Purdy RA, Kaufman DM, Langille DB. Comparison of problem-based learning and lecture format in continuing medical education on headache diagnosis and management. Med Educ 1998; 32:590-96.

- Elstein AS, Shulman LS, Sprafka SA. Medical Problem Solving. Cambridge, Mass: Harvard University Press; 1978.

- Eva, KW, Wood, TJ, Riddle, J, Touchie, C, Bordage, G. How clinical features are presented matters to weaker diagnosticians. Med Educ. 2010; 44: 775-85.

- Farmer EA, Hinchy J. Assessing general practice clinical decision-making skills: the key features approach. Aust Fam Physician 2005; 34:1059-61.

- Farmer, E.A., Page, G.　A Practical Guide to Assessing Clinical Decision-Making Skills using the Key Features approach.　Medical Education 2005; 39: 1188-1194.

- Fischer MR, Kopp V, Holzer M, Ruderich F, Junger J. A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities. Med Teach 2005; 27:450-55.

- Hatala R, Norman GR. Adapting the key feature examination for a clinical clerkship. Med Educ. 2002; 36: 160-65.

- Jacques A, Sindon A, Bourque A, Bordage G, Ferland JJ. Structured Oral Interview. One Way to Identify Family Physicians' Educational Needs. Can Fam Phys 1995; 41:1346-1352.

- Miller F, Jacques A, Brailovsky C, Sindon A, Bordage G. When to Recommend Compulsory versus Optional CME Programs? A Study to Establish Criteria. Academic Medicine. 1997; 72:760-764

- Norcini JJ. Setting standards on educational tests. Med Educ 2003; 37:464-49.

- Norman, G., et al. A Review of Recent Innovation in Assessment. In: Directions in Clinical Assessment. Report of the First Cambridge Conference on the Assessment of Clinical Competence, ed, R. Wakeford. Cambridge England: Office of the Regius Professor of Physic, Cambridge University School of Clinical Medicine, Addenbrooke's Hospital, Cambridge England. 1985: 8-27.

- Norman, G., Swanson, D., Case, S. Conceptual and Methodological Issues in Studies Comparing Assessment Formats. Teaching and Learning in Medicine 1996;8: 208-216.

- Norman, G., Bordage, G., Page, G., Keane, D. How Specific is Case Specificity? Med Educ., 2006; 40:618-23.

- Page, G. & Bordage, G. The Medical Council of Canada's Key Feature Project: A More Valid Written Exam. of Clinical Decision-making Skills. Acad. Med., 1995, 70: 104-110.

- Page, G., Bordage, G., Allen, T. Developing Key-Feature Problems and Examinations to Assess Clinical Decision-making Skills. Acad. Med., 1995, 70: 194-201.

- Page G, Boulais AP, Blackmore D, Dauphinee D. Justifying the Use of Short Answer Questions in the KF Problems of the MCCC's Qualifying Exam. In: Proceedings of the 9th Ottawa Conference, Cape Town. 2000.

- Sturmberg JP, Crowe P, Hughes C. Computer-assisted instruction: Guiding learning through a key features approach in a community-based medical course. Med Teach 2003; 25:332-35

- Tamblyn R, Abrahamowicz M, Dauphinee D, Wenghofer E, Jacques A, Klass D, Smee S, Blackmore D, Winslade N, Girard N, Du Berger R, Bartman I, Buckeridge DL, Hanley JA. Physician Scores on a National Clinical Skills Examination as Predictors of Complaints to Medical Regulatory Authorities. JAMA. 2007; 298:993-1001.

- Trudel, J., Bordage, G. & Downing, S. Reliability and validity of Key Feature Cases for the Self-assessment of Colon and Rectal Surgeons. Ann Surg. 2008 Aug; 248(2):252-8.