ITEM WRITING MANUAL FOR

MULTIPLE-CHOICE QUESTIONS

SAUDI COMMISSION FOR HEALTH SPECIALTIES
ITEM WRITING AND STYLE GUIDE

2011

1ST EDITION

Contents

INTRODUCTION

Multiple-choice questions (MCQs) are the most widely used test format in the health sciences today. Tests using MCQs are objective, can be machine marked and easily adapted for computer delivery. Most educational psychologists agree that this test format can test higher cognition, which is a necessity for the assessment of health care professionals. Over the last thirty years, there has been an increasing focus on the quality of examinations for professional licensing and certification. This in turn has led to the development of detailed guidelines, rigorous training of item writers and the application of strict criteria for quality assurance.

This manual and style guide will address some of these issues and in particular focus on the desirable standards that shall always be applied to Saudi Medical Licensing Examinations (SMLE) delivered by the Saudi Commission for Health Specialties (SCHS).

The manual is expected to be updated from time to time. The development is also expected to come from research carried out by SCHS using data from the testing activities that it is responsible for. It is not expected that what is contained in this manual can be applied to the universe of testing, but is born out of best practice for the purpose that SCHS will be using their testing services for. However, the text will always attempt to include a justification for the policies and guidelines presented so that the reader can follow the logic and arrive at the same goals that the authors of this document have.

Please note that the examples given are only examples used to illustrate a point or concept, and can always be improved on. Several experienced item writers and clinicians have read through the manual and checked the authenticity of the content including the content of the items used and their assistance is freely acknowledged at the bottom of this page. Finally, and importantly, no correct key is indicated as the examples have not been independently reviewed or used in an SCHS exam.

PREFACE

For those who may be less familiar with multiple-choice question formats and test measurement, this preface shall attempt to orientate the reader. There is also a glossary at the end of this booklet, Appendix E. [1]

A multiple-choice question consists of a stem with a question line (also known as the lead in question) at its end or underneath it. There is often a vignette or scenario embedded in the stem. SCHS will separate a vignette in the stem from the question line by placing the question line in its own paragraph, see diagram below.

The stem, with or without a vignette or scenario

Test Name	Test result with units	Normal values (with units)

Question line in separate paragraph

Options and distractors
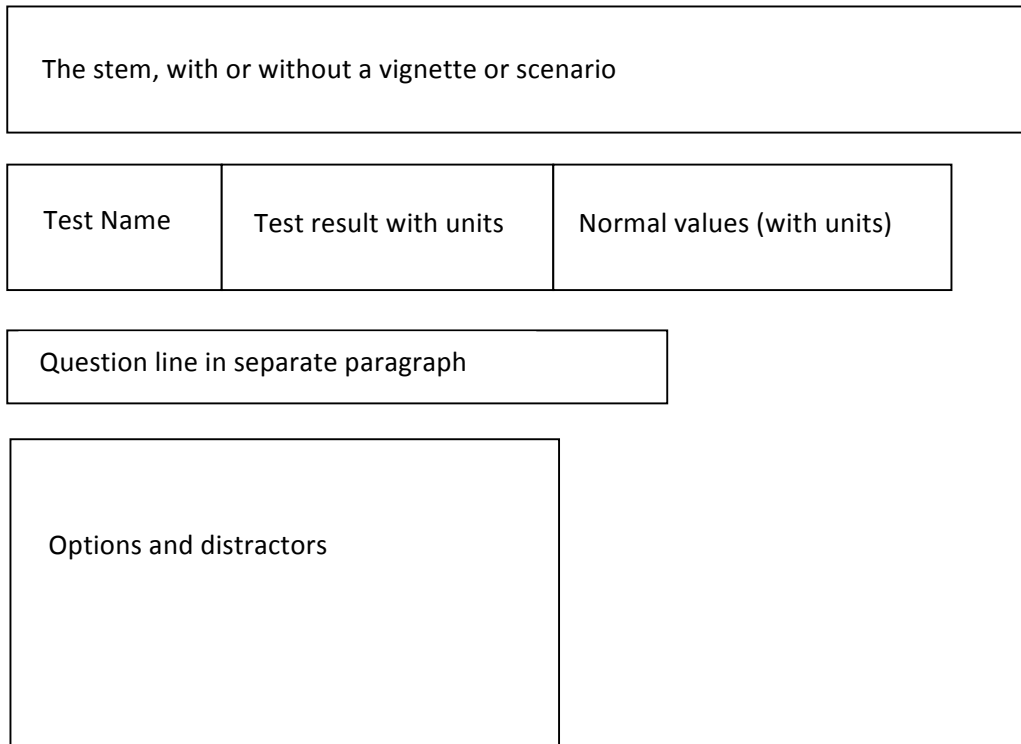
In this manual there are many examples of the above, and all MCQs used for SCHS licensing examinations shall be based on a clinical vignette. A vignette or clinical scenario may have laboratory results, in which case the results will be put in a table beneath the clinical vignette, see above, and note that units are always provided.

The number of options in an question can vary, SCHS have chosen to use four; but theoretically up to twenty-six may be used (utilizing all the letters in the alphabet). Many high stakes examination use five options because it is thought that five gives the correct *impression*, while making no difference to the way the question performs. One of the options will be the best choice, known as the correct key, while the others are described as distractors. An essential characteristic of all distractors, in the best one of four options for example, is that all options shall present plausible answers and if possible none shall be incorrect, as these are easily spotted as *filler[2]s* and quickly discarded, making the question easier. All options should be on the same continuum, for example all diagnoses or all forms of management.

---

[1] Glossary has been copied from www.pearsonassessments.com/pai/ai/research/resources/Glossary.htm. last visited on 11/17/2011.
[2] A filler is an option that is inadequate and seemingly has no place in the question.

Different types of multiple-choice questions have been designated with letters of the alphabet, which do not have any test measurement significance. The type that the Commission uses is the A-type (one best of four options). A popular type used until very recently in the UK was the X-type, with multiple true and false statements. This is not suitable for assessments in the health sciences as it leads to superficial understanding, and by no means can test how a candidate will use their knowledge.

Test measurements follow some important principles which need to be understood: such as item[3], validity, reliability, error of measurement, distractor functionality, item writing flaw, construct irrelevant variance, discrimination index (high low), p-value (or Difficulty), biserial and point biserial, corrected biserials, test blueprint, alignment and standard setting.

Validity and Reliability

Both these terms refer to a whole test, while validity also applies to a single item used in an assessment. Validity implies that the item or items used test what they are supposed to test. In a simple sense an MCQ item can test knowledge, the application of knowledge and analysis and or the synthesis of information. Clearly if an MCQ or a test paper with MCQs only seek to test recall of isolated facts, it can never test application of knowledge. Therefore, if one is claiming to test application of knowledge and uses such recall type questions this test is not a valid test of what is being tested for.

In order for an assessment to be truly valid it must be reliable or reproducible. In other words every item needs to be testing the same characteristic or trait. When every item tests the same trait and we have enough items, we take the candidate's scores or performance for the first half of all items and correlate these with the second half of items, a correlation coefficient is derived. Usually ≥ 0.85 is considered satisfactory for an institutional exam, while a national licensing examination shall have a reliability of ≥ 0.90. Reliability is calculated by a formula (e.g., Kuder-Richardson) or correlations derived using comparisons of halves (as above, or odd numbered items versus even).

Validity is determined by human judgment. And, validity is not necessary for reliability. It can be noted that for an examination to be reliable there shall be at least 100 items in the test. If on the other hand fewer items are used a formula (Spearman-Brown) will indicate for a given reliability how many items are needed to meet the required level.

Error of Measurement

All tests have measurement error. The more discriminating a test is, the greater is the measurement error, and so on one hand a desirable outcome is obtained and on the other the actual score given to a candidate becomes less secure. The error measurement becomes greater in proportion to changes in the standard deviation about the mean of all candidate scores.

---

[3] Item is a single test element, which might be a multiple-choice question.

Distractor Functionality

The point of using distractors is that the weak student will believe that they represent plausible answers to the question asked. However, if less than 5% of all candidates select a distractor, that distractor is said to be non-functional. The functional distractor frequency (FDF) even with five or six distractors is approximately 1.8-2.0 distractor per item. Clearly if the items are too easy or difficult (average correct responses above 90% or below 20%) then the number of functioning distractors is relatively less, the more desirable being when average test scores for the class are between 45% and 55%, see Fig 1. The item FDF is an indication of quality for an item, provided other data are considered. However, scores below forty percent cannot be said to demonstrate a sufficient amount of knowledge or its application to satisfy a professional examination.
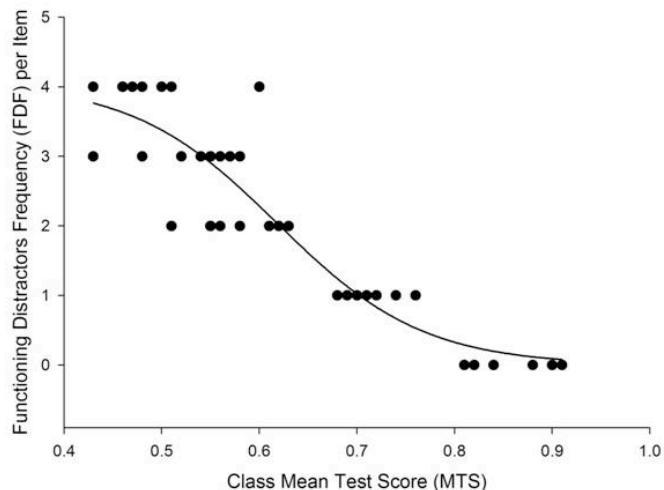

Fig 1
The mean test scores from a series of examination were used (n=35) and correlated with the mean FDF for each. Note, above 90% the FDF is zero and similarly below a MTS = 40% all four distractors (A-type, one best of best options) are functional; Ware and Mohammed, 2010.

Item Writing Flaws (IWFs) and Construct Irrelevant Variance

There are some generally accepted guidelines about item writing but only recently in health sciences was it possible to demonstrate that the presence of item writing flaws can lead to incorrect exam results. Many believe that writing negatively phrased questions is acceptable because the item tests a wider range of knowledge. The wider range is only tested at the level of recall or memorized fact, while tests in health sciences need application, analysis and synthesis. Tarrant and Ware have demonstrated that using examinations with a substantial number of flawed items leads to passing a significant number of borderline candidates who should not have passed and denying an even greater number of distinction candidates their true rewards (Medical Education 2009). With proper preparation and training, the frequency of IWFs shall be $\leq$7% in any assessment (Ware and Vik, 2009). It is extremely important that any reviewer is familiar with recognizing IWFs and aware how best to edit them out without changing the focus or meaning of an item.

Borderline candidates often depend on using IWFs and test wiseness to gain extra marks without having done the work necessary in the course studied. Therefore, for these candidates the assessment not only tests their learnt

knowledge but also their test taking ability and use of test wiseness, two attributes which distort the score awarded. This false measurement is called construct (the intended focus of the test) irrelevant variance. Another example might be when jargon or poor language causes the candidate to be confused, giving rise to the often complained of problem with MCQs, when the candidate does not know exactly what the question is being asked for.

High-Low Discrimination Index (DI)

A good MCQ item shall discriminate between a top student and weak student. The top student will be able to score more correct items than the weak student. So, if one takes the proportion of the top students who get a given item correct and those from the bottom students and then determines the numerical difference, figures between +1 through zero to -1 are obtained for every item. Note that if the MTSs for the assessments are either high or low (cf., FDFs) the discrimination indices will be near to zero or even negative. A good DI is ≥ 0.4 and anything below 0.10 should be revised, while DIs at zero or negative are unacceptable and the item should usually be revised. DIs are useful markers for quality (Ware and Vik, 2009).

Difficulty, (facility) or p-value

All three refer to the proportion of candidates that marked the correct key for a given item. Ideally a test shall have an MTS of 50% - 70%, as this will give the highest mean DI and reliability.

Biserials (including point and corrected)

Point biserial coefficients (RPBS, apply only to MCQs), biserials (RBS) and corrected RPBS and RBS are correlations between candidate performance on any item and their overall performance on the whole test. Just as negative high-low discrimination is an indicator of a poor quality or flawed item, a negative biserial is an even stronger indicator of the same. Usually such questions are removed before calculating the final score for each candidate. A corrected biserial is the same index but the item in question has been removed before calculating the final result for comparison with the actual item.

Test Blueprint (TB) and Alignment

For any assessment to represent a valid test of whatever it is that is being tested for, the content of the examination shall be strictly controlled. This is partially achieved by writing down exactly what shall be tested and having a panel of experts agree on this. Usually the TB is drawn up as matrix, because many skills are common to a large part of content, for example making a diagnosis. Usually the TB represents an alignment between the test items themselves and what has been studied or the expected outcomes should be . For postgraduate or licensing examinations it is also very important to publish these expectations *up front* as a series of objectives or outcomes.

Standard Setting

Depending what purpose an examination serves will determine the type of standard that is set. Usually licensing and certifying examinations will use what are called absolute standards, which means the standard is set by a regulatory authority (university of the Saudi Commission) or chosen by a series of judges using one of the acceptable and well researched methods.

Examinations which serve the purpose to select candidates who are best suited for a course or training, usually are standard set with a relative methodology. This means the candidate results are ranked and the top candidates are selected.

There are almost thirty methods used for standard setting, but at the moment an absolute standard of 60% is used for most SCHS examinations.

BACKGROUND

Health care professionals shall receive a license to practice because they have sufficient and appropriate levels of knowledge, skills and attitude. Multiple-choice tests[4], however, can only test the knowledge component. In order that this question format can contribute to the testing of professional competence, the item writer should always frame any knowledge sought for in a meaningful context. This is almost invariably achieved by using vignettes or scenarios. **Q1** shows a typical vignette based MCQ.

Q1.
A typical example of an MCQ with a vignette satisfying the hand cover test. Compare **Q2** which does not, but in that form is acceptable.

A 58 year-old man had epigastric pain radiating into his back for three months and now came to see his family doctor. The man has lost eight kilos in weight during the same period and on examination appeared anorexic, jaundiced and pale. The doctor was very concerned and arranged a referral for the patient to the hospital and an urgent specialist's opinion.

What was the most likely diagnosis?

   A.   Chronic pancreatitis
   B.   Obstructive jaundice
   C.   Viral hepatitis
   D.   Pancreatic carcinoma

Writing questions like the one shown in **Q1** has certain advantages apart from satisfying the requirements of a licensing examination. Having written the vignette it is possible to derive a whole series of other relevant questions related to the same patient context, see **Qs 2-3**. The three questions use the same clinical context, but emphasize different and important issues. Moreover, if the three questions were used in the same examination, the information in one would serve to help answer one or more of the others. For example, the issue of priority being asked for in **Q3**, is answered by the last sentence in **Q1**, which was removed in **Q3**.

If the first example with the same context takes 20-30 minutes to write, each related question might only take 5-10 minutes to complete. Therefore, the use of vignettes can have great advantages when constructing MCQs.

Factual knowledge related to this particular context could have been asked for, but provides no insight into how the candidate might use their knowledge. **Q4** and **Q5** are examples of recall knowledge related to the same clinical area as in **Qs 1–3**. It is clear that candidates correctly answering any **Qs 1–3** have demonstrated more useful knowledge than candidates who correctly answer **Qs 4–5**.

---

[4] In this manual multiple-choice questions shall be referred to as "MCQs" and are all A-types, or best one of four. SCHS has a policy to use four options, of which one will be the correct key.

Q2.
Again using the same vignette as in **Q1**, the question asked is modified to include only those diagnoses in the option list, and now a different emphasis is achieved with the same context.

A 58 year-old man had epigastric pain radiating into his back for three months and now came to see his family doctor. The man has lost eight kilos in weight during the same period and on examination appeared anorexic, jaundiced and pale. The doctor was very concerned and arranged a referral for the patient to the hospital and an urgent specialist's opinion.

Which of the following is the most likely diagnosis?

A. Chronic pancreatitis
B. Metastatic disease
C. Obstructive jaundice
D. Viral hepatitis

Q3.
Again using the same vignette as in **Q1**, with the removal of the final sentence, another question is asked, and more could be.

A 58 year-old man had epigastric pain radiating into his back for three months and now came to see his family doctor. The man has lost eight kilos in weight during the same period and on examination appeared anorexic, jaundiced and pale.

What referral priority should be given?

A. < one month
B. $\leq$ one week
C. Same day
D. Emergency transport

Q4.
An illustration of a recall type question related to the context, pancreatic carcinoma, used for **Qs 1–3**.

Which organ is epigastric pain radiating through to the back most likely coming from?

A. Gallbladder
B. Right kidney
C. Liver
D. Pancreas

Q5.
Another example of a recall question related to those in **Qs 1–3**.

Which two clinical features best characterise pancreatic carcinoma?

A. Epigastric mass and jaundice
B. Epigastric pain relieved by sitting forwards
C. Pain radiating to back and weight loss
D. Large volume vomiting and epigastric mass

Most of the examinations used for licensing by the SCHS are screening tests to exclude unsuitable doctors from practice of their specialty in Saudi Arabia. They shall test wide and representative samples of the specialist's knowledge and competence base. In this respect, applied knowledge is considered an important measure of competence. Vignettes set the context for a test of applied knowledge, reasoning, analysis and interpretation. This is at the very core of examining professionals for the granting of a medical or any other form of professional license.

Merely testing recall of factual knowledge is not the goal of an SCHS screening test to grant a medical license.

Finally, all test items shall be aligned with the purpose (SCHS Test Blueprint) / curriculum / course / training that it is examining. The final arbiter being objectives: such as: *Explain the meaning of beneficence, its implications and usage with common ethical problems*. **Q6**, illustrates an item that is aligned with this objective.

Q6.
An example of a question that is aligned with an objective related to beneficence and ethics.

A 15 year-old girl with anorexia nervosa has been attending her school clinic for one year, but still continues to lose weight. Now her parents wish her to be seen by a specialist but first wish to talk with the school nurse. The nurse is initially uncertain what to do.

What should guide her decision?

- A. Confidentiality
- B. Respect and autonomy
- C. Veracity
- D. Beneficence

Above is an illustration of test alignment with a learning objective, however, it is to be remembered that SCHS items shall be aligned with a Test Blueprint cell (see Appendix **A**) and topic list provided.

LANGUAGE, PRESENTATION FORMAT, VIGNETTES AND WORD COUNTS

Language is an important issue for SCHS, as it does not wish to present assessments where language competence represents more than a minimum part of the test. This can be avoided if attention is paid to the language used. There are three considerations to make an MCQ easily and quickly understood. The use of the three "Ss" may be helpful to remember

1. Structured format
2. Simple language
3. Sequence logic

Structured Format

The SCHS's in-house style of MCQs follows a structured format consisting of a relevant vignette, tabulated vital signs and lab results, if any; followed by a clear question and a list of four options. A structured format allows item writers to write their test items so that they fulfill two criteria: (a) follow the in-house style and (b) can be easily uploaded into a database by adhering to a strict template format. Three questions are given as examples with increasing amounts of information, **Q 7-9**.

Example **Q7**, has the simplest format. It uses no vignette and tests the lowest level of cognition. The question asked is clear but only testing recall of isolated fact. **Q8** has a vignette with the clinical features of the patient's presentation, although not all (the result of an examination of the neck has been omitted). However the vital signs recorded are given and provide important further information about the patient. The presentation format using tabulation makes these data quick and easy to understand just by scanning the lines. This item tests higher cognition.

Q9 is a more complex MCQ with both lab and measurement data added to the original vignette. Importantly when lab data are added normal ranges are given and the values shall always be in SI units. The wpi (words per item) for **Q9** is 111, of which 96 is accounted for by the vignette.

Q7.
An example of a simplistic type of MCQ, testing only recall.

> What condition is associated with dry thin hair, bradycardia and slow relaxing reflexes?
>
> A. Renal failure
> B. Hypercalcaemia
> C. Addison's disease
> D. Myxoedema

Q8.

Physiological measurements added in a tabulated format to provide the reader with easily read information.

A 68 year-old woman was referred to the clinic with gradual loss of memory, generalised muscle weakness and difficulty hearing. On examination her mood was depressed, she had pale conjunctivae, proximal muscle weakness and impaired hearing. The vital signs recorded were:

Heart rate	60 /min
Respiratory rate	16 /min
Blood pressure	170/90 mmHg
Temperature	38.0 ^{o}C

What was the most likely diagnosis?

    A. Anxiety
    B. Multiple sclerosis
    C. Dementia
    D. Myxoedema

Q9.

An example of a more complex MCQ with both clinical measurements and lab data given. The quantity of information to take in requires that the presentation shall be so clear that a single scan is sufficient.

A 68 year-old woman was referred to the clinic with gradual loss of memory, generalised muscle weakness and difficulty hearing. On examination her mood was depressed, she had pale conjunctivae, proximal muscle weakness and impaired hearing. Her vital signs and relevant lab results are given below:

Heart rate	60 /min
Respiratory rate	16 /min
Blood pressure	170/90 mmHg
Temperature	38.0 ^{o}C

Test	Results	Normal values
TSH	10.1	0.4-5.0 mU/L
Free T$_3$ (free)	3.6	3.5-6.5 pmol/L
TPO antibody	4356	10-30 IU/L
Hb	90	120-160 g/L
MCHC	380	320-360 g/L
MCV	9100	86-98 fL

What was the most likely diagnosis?

    A. Hypopituitarism
    B. Renal failure
    C. Dementia
    D. Myxoedema

Simple Language

One rule for the language level used in assessments presented to candidates who are non-native English language speakers, is the use of *simple language* and avoiding making the questions and or the sentences within them, too long or complex. The word counts for **Qs 1–3** are 45–65 wpi. **Q10** is a representative example of the complex (linguistically) format used by the National Board of Medical Examiners (NBME) in a USMLE (United States Medical Licensing Examination) Step 2 examination.

The example in **Q10** has a wpi of 241, and the vignette alone has 211 words. This is entirely reasonable for licensing a doctor who will practice in the United States and use American English as the language for communication. This is not satisfactory where English is being used to license a doctor by SCHS and who will usually be using the Arabic language, but also need to be comfortable communicating in English often with colleagues for whom English remains a second or even third language.

Q10.

An example of a multiple-choice uestion used in USMLE Step2.

> A 32 year-old Bedouin woman has had ten years of gradually deteriorating attacks of bronchospasm for which her family practitioner has prescribed a short acting beta agonist delivered by a metered inhaler. During the last couple of months the patient has given almost twice as many doses per day compared with earlier. Previously, her tight chest was well controlled with glucocorticoids inhalation. Mona has become increasingly anxious because of a family feud for which she has been given a selective serotonin reuptake inhibitor (SSRI). On top of all of this she has also suffered from migraines for which she was quite successfully treated with the sulphonamide derivative, sumatriptan. As a result of all the stress her migraines became worse and her own doctor added a beta blocker to her drug list. She also takes an oral contraceptive as she already has six children.
>
> During the last three months while all the stress was developing she has taken to drinking coffee many times each day, compared to her usual cup of green tea after breakfast only. She is now becoming increasingly tired and worn out and has now come to see her doctor who finds, on examination, that she has widespread and marked expiratory wheezes, finding no other abnormality on his examination.
>
> Which of the following is most likely to have been responsible for the examination finding?
>
> A. Beta agonist
> B. Caffeine intake
> C. Oral contraception
> D. Beta blocker
> E. The SSRI

Using the example in **Q10** the same content can be examined using a more simplified language as seen in **Q11**. The wpi has been reduced from 241 to 88, with no loss of information gained from the candidate who marks the correct keyed option. **Q11** does not test the same language ability as **Q10**. Normally, measuring language ability in professional knowledge tests would be considered an example of construct irrelevant variance (Downing SM, Haladyna TM, 2005) and undesirable. However, the NBME are very conscious of communication problems being a major cause of malpractice suits in the United States and believe that a measure of clinical competence shall include the language used for communication.

Q12 tests the same content as **Q6**, and demonstrates an example how choice of words can cause problems for non-native English language speakers. Although there has been a small increase in words, from 67 to 85 wpi, the language complexity has increased relatively more.

Q11.
An example how stringent editing can remove issues of language to a large extent from an item, compare this question with **Q10**.

> A 32 year-old female diagnosed with an anxiety disorder is being treated with fluoxetrine. She also has asthma previously well controlled with glucocorticoid inhalations, and migraine treated with sumatriptan to which has recently been added propranolol for better control. She takes an oral contraceptive and has been drinking more coffee over the last two months. Now, she must use her albuterol inhaler more frequently.
>
> What is causing this increased use of her inhaler?
>
> A. Beta agonist
> B. Caffeine intake
> C. Oral contraception
> D. Beta blocker
> E. The SSRI

Q12.
The language in the vignette and question line will present a non-native English language speaker with problems, compare this question with **Q6**.

> A tenth grader is feeling overwhelmed by her dysfunctional family antics. She has vented her inner angst through an eating disorder. Despite the unwavering support from a school clinic care giver she continues on the downward spiral of weight loss. Now her parents wish her to see a specialist but are persuaded by their offspring to first allow her clinic support to speak on her behalf.
>
> What should guide the care giver's decision?
>
> A. Confidentiality
> B. Respect and autonomy
> C. Veracity
> D. Beneficence

It is important to ensure that items used by SCHS fit with the in-house style and policies and whether the items falls within the prescribed wpi, which can be stated as:

A complex vignette, as in **Q9**, shall not exceed 100 words.

A simple vignette, as in **Q11**, shall not exceed 70 words.

(use Microsoft$^{©}$ Word Review function to count the words)

It is not possible to prescribe for the lead-in question line plus options, but they shall be kept short and precise. To achieve a reasonable wpi several other rules shall apply and are illustrated below:

Repetition of references or units in every option

Q13 illustrates two important style issues: overlap in ranges and repeated terms. When ranges are given there shall be no overlap. In this example, option **A** overlaps with option **B**, option **B** with option **C**, and option **C** with option **D**. Repeating reference words (*of blood volume*) is unnecessary and is unacceptable. Instead, **Q14** shows how to re-word the MCQ and includes one more piece of vital information by setting the context with the statement that the blood loss shall be related to a normal adult.

Q13.
An MCQ with overlapping ranges and repeated reference words in every option that are unnecessary, see **Q14** for the solution.

> How much blood loss will cause irreversible organ damage?
>
>   A.   5-10% of blood volume
>   B.   10-20% of blood volume
>   C.   20-30% of blood volume
>   D.   30-40% of blood volume

Q14.
Q13 has been edited to improve clarity and also remove ambiguity by setting the conditions, or context, for the loss of blood.

> What percentage BV loss will cause irreversible organ damage in a normal adult?
>
>   A.   5-10%
>   B.   11-20%
>   C.   21-30%
>   D.   31-40%

In **Q14** an abbreviation for blood volume (BV) has been used. This should be spelt out. Despite that units shall be given in tables for measurements and laboratory values, they don't need to be repeated in every option, given that the values are in the same units, instead the units specified shall be stated in the lead-in question line, see **Qs 15** and **16** for examples of this practice. Similarly, reference for every patient A-D is unnecessary.

Q15.
Another example of repeating unnecessary units and references, see **Q16** for the edits.

> Below are the blood pressures taken from five adult patients (A-D) attending a family practitioner's clinic.
>
> Patient A   180/11 mmHg
> Patient B   160/90 mmHg
> Patient C   144/64vmmHg
> Patient D   130/88 mmHg
> Patient E   116/78 mmHg
>
> Which patient is pre-hypertensive?
>
> A. Patient A
> B. Patient B
> C. Patient C
> D. Patient D

Q16.
Q15 has been edited to remove all unnecessary repetitions.

> Four measurements of blood pressure in mmHg were taken from four adult patients (A-D) attending a family doctor's clinic.
>
> Which patient is pre-hypertensive?
>
> A. 180/110
> B. 160/90
> C. 144/64
> D. 130/88

Exams delivered by SCHS are not *speeded tests*. In other words, use of unnecessary words shall be avoided. It is also possible to use what may be called abbreviated MCQ language, whereby definite and indefinite articles will often be omitted. In general, non-English language speakers will take approximately one minute to read one hundred words, Figure 1.
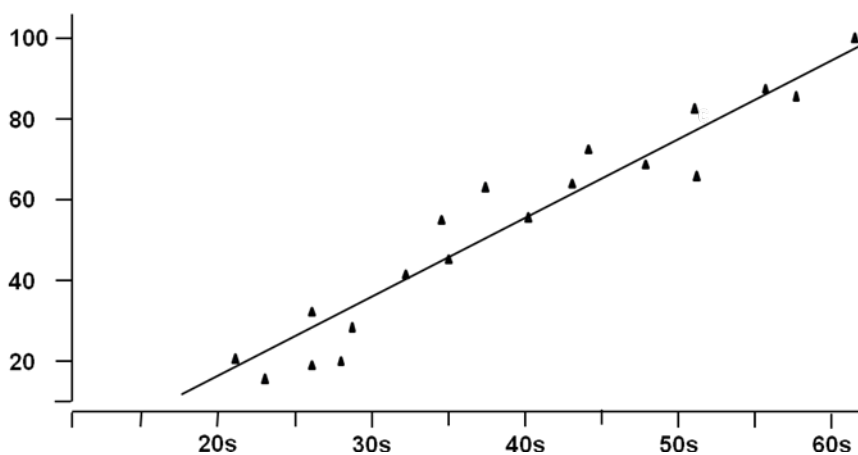

Figure 1. Items with varying wpi read to comprehension, $r^2 = 0.95$, showing that the expected examinees will read approximately one hundred words per minute. The x-axis is time taken to read the item in seconds and y-axis the number of words (wpi) in the item.

<u>Abbreviations and Acronyms</u>

All abbreviations and acronyms shall be spelt out on first use, except when used in options to avoid length disproportions when a foot note under the item can spell out the meaning. What seems natural for a native English language speaker may not be for a competent doctor for whom English is their second or third language.

<u>Colloquial English</u>

Colloquial (informal) and fluent language is another problem that may be faced by non-native English speakers and examinees from different cultures. Several examples are given to emphasize and explain the problem:

In Africa *a thirsty maid might take a soda in desperation*, while *an Indian maid who took soda in desperation* was trying to commit suicide. The English might offer each other *soft drinks*, a Scotsman a *ginger*, a Londoner *pop* and a New Yorker a *tonic*. Care should be taken before assuming that the language was clear to all readers.

Instead of *a patient who has gotten ill after eating a salad*, it is better to write a patient who became ill after eating a salad. *A patient's memory was poor* instead of *the patient had frequent lapses of memory*, or *he had excruciating pain* should be changed to *he had severe pain*. It may appeal to item writers to insert humour into their questions, such as*: Jack pricked up his ears when he heard that the old man was over the hill but was still taken aback when he kicked the bucket*. Instead, Jack was surprised to hear that the old man had died. Often the only way to identify and remove one's own colloquialisms is to read what one has written after a pause of several days. By doing so the word blindness we all have seems to *melt* away!

Sequence logic

The last of the three "Ss" is an important sequence that must be followed if the order in which patient information presented in a vignette is to be easily assimilated and understood. The sequence is based on the way we write our clinical notes and it is important that the data is not given out of order. However, an MCQ vignette is not a complete set of clinical notes and therefore the information shall be selective. Patients in MCQs do not need names or gender unless it is important for clarification of fact; therefore, MCQ vignettes are stripped-down clinical notes. If relevant the following may be included:

> Age
> Gender
> Ethnicity / race or culture (if not stated the dominant form in the country)
> Presenting complaint
> Relevant patient histories without medical or jargon terms
> Selected findings of a physical examination (if relevant)
> Medication being taken (with dose details only if relevant)
> Tabulated vital signs
> Tabulated lab data with normal values
> Other relevant test results (e.g. blood smear or CT scan)

The vignette in **Q9** was a good example of the sequence given above.

QUESTION LINES, ACCEPTABLE VARIANTS

Clearly stated questions shall be used to the exclusion of sentence completions. The reason for this policy is to increase the clarity of every item and as far as possible to remove ambiguity. **Q17** and **Q18** illustrate an example of a sentence completion edited to become a clearly stated question.

Q17.
The lead-in question uses the sentence completion format, which can easily be re-written with a clearly stated question line, see **Q18**.

A car driver was involved in a head-on collision and received a severe blow to his chest over the sternum. He was admitted to the hospital and during the first survey of his condition the doctor noticed that he has a markedly raised JVP.

These clinical features are most likely explained by:

A.   Pulmonary embolism
B.   Acute heart failure
C.   Atrial fibrillation
D.   Pericardial effusion

Q18.
The edited version asks a question that can be answered without reference to the four options below it. In **Q17** it is necessary to review the four options below the sentence completion to choose the answer that is most appropriate from the listing.

A car driver was involved in a head-on collision and received a severe blow to his chest over the sternum. He was admitted to the hospital and during the first survey of his condition the doctor noticed that he has a markedly raised JVP.

What is the most likely cause?

A.   Pulmonary embolism
B.   Acute heart failure
C.   Atrial fibrillation
D.   Pericardial effusion

If the correct key is not the most commonly recorded best answer that would be given by an expert, without reference to the option list, the lead in question must state that only the choices in the option listing shall be used and in which case the option list shall not include the expert's best answer but clearly a best choice, see **Q19**. This type of question format shall only be used to test alternative applications for the same knowledge or competence. Age or gender may therefore become a factor which influences the choice, and if so must be stated.

Most of the time, the best answer shall be asked for, in other words the commonest cause, diagnosis, explanation, dosage form or diagnostic investigation used. Unambiguous language clarity is essential, and ambiguity must be removed from any question line. **Q20** is an example of a potentially ambiguous question. The patient must surely receive an intravenous diuretic, preferably as soon as possible, but he will also benefit from high flow oxygen, so which comes first?

Q19.
An example of an item that may not be presenting the most often recorded cause, instead asks the question which is the most likely from the options provided.

> A cardiologist was examining an 18 year-old patient. First, with his right hand on the patient's left chest he felt a *parasternal heave* and with a stethoscope heard a splitting of the second heart sound.
>
> Which of the following is the most likely cause?
>
> A. Aortic stenosis
> B. Mitral regurgitation
> C. Congestive heart failure
> D. Patent ductus arteriosus

For the correct key to be **C** and not **A**, all the experts must agree, but this will not always be the case. When experts cannot agree, the item will need editing until there is agreement or rejection. The problem lies in the use of the word *first* as the qualifier for what shall be done. Finally, while considering **Q20**, it can be stressed that this is not a suitable test of knowledge for any SCHS exam. Rather, refer to **Q21**, where a vignette is used to good purpose.

Q20.
An example where what is being asked for may not be obvious to an exam candidate. This typically happens when question lines are formulated like this one.

> A man presented to the Emergency Department with acute onset pulmonary oedema.
>
> What should be done **first** for this patient?
>
> A. Intravenous diuretic administration
> B. Endotracheal intubation
> C. High flow O_2 at 8L/min
> D. Immediate admission to ICU

Q21.
The vignette makes this item potentially useful, however, the ambiguity still remains and will need a further edit as in **Q22**.

> A 25 year-old man suddenly experienced acute left sided chest pain and shortly afterwards became increasingly breathless. He was brought to the Emergency Room in a state of collapse
>
> What should be done first for this patient?
>
> A. O_2 at 8L/min by face mask
> B. Endotracheal intubation
> C. Insertion of a chest drain
> D. Immediate portable chest X-ray

Although the likelihood of **A**, **B**, **C** or **D** being the best answer is seemingly similar to Q20, which is considered unacceptable, item **Q21** has now been given a context with sufficient clinical detail to make the item test higher cognition. But it still does not have perfect clarity.

There is still a problem with **Q21**, as it is debatable which action should be taken first, **A** or **C**, although logically it should be **A** as it is the easiest to take. Simple editing can make **Q21** unambiguous, see **Q22**. Now, option **C** can be the only correct key, although the question line would need to be worded another way: *What is the definitive action for this condition?* Or, asked as in **Q22**.

Q22.
The ambiguity in **Q21** is now removed by asking for a definitive (corrective or curative) action, not just supportive. High flow oxygen would be considered supportive in this case.

> A 25 year-old man suddenly experienced acute left sided chest pain and shortly afterwards became increasingly breathless. He was brought to the Emergency Room in a state of collapse
>
> What definitive action must be taken?
>
> A. O_2 at 8L/min by face mask
> B. Endotracheal intubation
> C. Insertion of a chest drain
> D. Immediate portable chest X-ray

Not all lead-in question lines are clear, below are four examples not to be used followed by four example items:

1. What is the first line of treatment or management?
2. What will be the next step in management?
3. What will be the final investigation?
4. Which of the following should be done next?

Q23.
When a question uses "first line of management" it almost implies a protocol driven choice, and maybe not the most commonly used or absolute best.

> On examination of a 65 year-old man, wide splitting of the second heart sound was found as an isolated finding in an asymptomatic patient.
>
> What is the first line of management?
>
> A. Give reassurance
> B. Prescribe digoxin
> C. Prescribe a β-blocker
> D. Implant a pacemaker

When considering the possibilities to answer **Q23**, it is as well to remember the guidance given by the instruction under the heading of management in the Test Blueprint (Appendix **A**):

> Management includes a plan of investigations, any specific medical treatment, non-medical aspects of management like surgical or psychosocial and or major decisions regarding admission (e.g. ward vs ICU, isolation, discharge).

In **Q23**, there are therapeutic, non-medical, surgical and physical interventions (option lines), all as possible choices, legitimately included under the same heading according to the Blueprint. However, heterogeneity should be avoided as the options are not the same form of management, i.e., different forms of management.

Q24.
A question asking for
the next step in
management.

> A 35 year-old patient presented with stabbing left chest pain and palpitations. On examination, a mid-systolic click was heard followed by a late systolic murmur.
>
> What will be the next step in management?
>
> A. β-blockers
> B. Referral to cardiologist
> C. Electrical cardioversion
> D. Echocardiogram

Q24 illustrates another ambiguous question line which should not be used and the options are not on the same continuum, they are heterogeneous.

Q25 is yet another item whose question line lacks clarity. The question asked has no obvious answer because of the wording for the question: "Which of the following should be done next?" Items used by SCHS in their licensing exams shall have issues of ambiguity removed by scrupulous editing.

Q25.
The question line asks
for the one (single)
investigation to be
done (ordered) next.
The choice is not clear.

> On general examination of a 35 year-old man the doctor noticed that he had a corneal arcus and the presence of Xanthelasma.
>
> Which of the following should be done next?
>
> A. Serum cholesterol
> B. Blood urea
> C. Fasting blood sugar
> D. Serum troponin I

THE OPTIONS, CORRECT KEY AND DISTRACTORS

SCHS policy is to use MCQs with four options, which will also save on item writer's time, increase test efficiency, meaning that more testing of every candidate's knowledge is possible in a given time and does not diminish the psychometric properties of the test or its items. All options shall be on the same continuum (Figure 1), i.e., all are either diagnoses, investigations, medications, etc.

Figure 1. The options continuum scale.

Q25 is an example of where options are on the same continuum. All four options are the same "sort", being lab tests. **Q24** is an example where options do not lie on the same continuum with all four options offering a different solution, therapeutic, a referral, active treatment and finally an investigation. The problem is that when they are not part of the same continuum the question can become ambiguous.

In addition to being on the same continuum, the options should satisfy the following criteria:

- All options shall be the same length (within two to three words if possible)
- Preferably, no longer than three quarters of a line in length
- Avoid mutually exclusive statements, not an absolute, but use sparingly
- Always consider dropping the definitive or indefinite articles (*the* and *an*)
- Always edit sentences to the shortest form to avoid getting each option too long
- Any commonality in every option move upwards into question line
- Keep choices homogeneous in content and grammatical structure.

Q26 is an example of options of perfect length while **Q27** is an example of an item with options that are unnecessarily long while **Q28** is the edited version. The main problem is the length of the options and the difficulty it creates for non-native English language speakers.

SCHS has the policy to limit the wpi by limiting the words in the vignette, 70 without tabulated data and 100 with. This will mean that the maximum wpi will be <150 even with a vignette with tabulated data. The question line with four options shall usually be about 30-35 words, in **Q27** the question line and options are 101 words.

Q26.
All options the same length, same continuum and less than three quarters of a line in length. Note the clarifications at the bottom in italics.

On examination of a patient's chest the following findings were recorded

Chest expansion	reduced
Percussion note	dull
Breath sounds	bronchial
Added sounds	crackles
Vocal resonance	increased

What was the most likely pathology?

A. Emphysema
B. RUL atelectasis
C. RML cyst
D. RUL pneumonia

RUL: Right upper lobe
RML: Right middle lobe

Q27.
Apart from being a long question, wpi =190, the vignette is almost half of the total wpi (89), leaving the options to be two or three lines each.

A 3 year-old child presented with five minute rage episodes in response to simple parental orders. The child is hyperactive, but does not interfere with other children. At bedtime she needs constant encouragement to make any progress towards bed. Her sleep is good with occasional bed wetting; weight and height at the upper 25th percentile with no past medical history. Developmental milestones and a full physical examination are normal. In the presence of adults the child appears shy, but becomes easily distracted and can play normally in their presence.

From the following which of the following options would now be the most reasonable step to take for this child's management?

A. Treating physician and social worker should sit with the child's parent and inform them this situation will improve by itself with time.
B. Play therapist should put a plan with the family regarding specific plays and games to build up certain social skills.
C. This is a clear pediatric depression presentation and has to be started on antidepressant medication.
D. This child is facing a problem in paying attention and the treating physician has to start him on methylphenidate.

Q28.
Editing bringing about
a substantial change in
item **Q27**, but tests the
same subject matter
and issue.

> A 3 year-old child presented with five minute rage episodes in response to simple parental orders. The child is hyperactive, but does not interfere with other children. At bedtime she needs constant encouragement to make any progress towards bed. Her sleep is good with occasional bed wetting; weight and height at the upper 25th percentile with no past medical history. Developmental milestones and a full physical examination are normal. In the presence of adults the child appears shy, but becomes easily distracted and can play normally in their presence.
>
> From the following which would now be the most reasonable step to take for this child's management?
>
> A. Reassurance
> B. Play therapy
> C. Trial of fluoxetine
> D. Trial of methylphenidate

Whereas **Q27** is difficult to read, **Q28** is easier and the content can probably be memorized after a single read through. The total wpi has been reduced from 190 to 119 and the number of words in the question line and options from 101 to 30. These differences are large and will not just improve a candidate's performance but prove to be less tiring because it does not require the same level of concentration to read through and understand the information.

Many poorly constructed MCQs have a series of mutually exclusive statements, **Q29**. The last option also has little to do with the first three and for most examinees it is an obvious filler, and ignoring it makes the MCQ easier to answer correctly. It has to be assumed that the candidates know that reduced resonance does not mean stony dullness. Once again, the vignette has been complicated by unnecessarily advanced language, although asthenic is classically used in many textbooks to describe individuals who seem more at risk from developing a spontaneous pneumothorax.

Q29.
Typical example of an
MCQ with mutually
exclusive options (**A**-**C**).
Note too, option **D** is a
filler, as it seems to be
of such a different
quality to the others.

> A rather asthenic boy suddenly developed left sided chest pain becoming increasingly breathless after. The Emergency Room doctor carried out an examination
>
> What might he have been found on percussion of the left chest?
>
> A. Increased resonance
> B. Decreased resonance
> C. Unchanged resonance
> D. Stony dullness

Why are MCQs with mutually exclusive options undesirable? In this case any normal percussion findings can be ruled out because there are three other options with abnormal findings on percussion, that leaves **A**, **B** and **D**. Based on convergence, the answer must be **A**, **B** or **C**, and with **C** ruled out there is a 50/50 chance it is **A** or **B**, much better than the original 25%.

Every MCQ, best one of four options, must have one satisfactory correct key. Although, it shall be stressed, that all good distractors must be plausible. When only one answer could possibly be correct the nature of the MCQ becomes true and false, which then requires recall of an isolated fact and thereby tests only lower cognition. Much preferable is an MCQ with all four options being potentially correct but one which is clearly a better answer than the others, see **Q30**.

Q30.
An example of good options that are all plausible (and two words each). All will give dullness to percussion, while only with a pleural effusion will there be stony dullness on percussion.

A patient who was known to have tuberculosis would have been expected to have a cough and possibly slight dullness to percussion at one or both apices of the lungs, but this patient had evidence of stony dullness at his right lung base.

What are these findings confirmation of?

A. Pneumonic consolidation
B. Liver enlargement
C. Pleural effusion
D. Pleural thickening

Q31.
This item has two answers without the result of a further investigation. With two mutually exclusive options it is also defective.

A patient who had his hearing tested underwent a Weber test. He said he heard the sound best on the left side, and nothing on the right.

What are these findings suggestive of?

A. Right sided conduction deafness
B. Right sided sensorineural deafness
C. Left sided sensorineural deafness
D. Left side conduction deafness

Weber's test lateralizes to the side with a conduction problem, while nothing is heard on the side with a sensorineural problem, therefore, there is no best answer to **Q31**, without the results of a Rinne's test and or a simple whispering test to determine if there is any problem at all. Moreover, double mutually exclusive option pairs are undesirable.

All good distractors are plausible choices for the question asked. However, the details of the question and the vignette refine the choice to the correct answer. **Q32** is an MCQ with good distractors. All the options are plausible, while option **C** is the best answer.

Q32.
An example of three distractors which are all reasonable for the facts of the case.

A 55 year-old miner and heavy smoker with a chronic cough, was admitted to the medical ward for evaluation. The Intern examined his respiratory system and was not surprised to find a chest hyper-resonant on percussion.

What is the most likely cause of this finding?

A. Chronic bronchitis
B. Restrictive lung disease
C. Emphysema
D. Chronic obstructive lung disease

THE CONTENT

Every exam shall be created with a Test Blueprint (TB) supported by a topic list, see examples in Appendix A. The actual TB has abbreviated headings and the list below gives the subheadings[5]:

Diagnosis including: clinical reasoning, interpretation of symptoms and/or signs to reach the most likely diagnosis, proper interpretation of investigations.

Management including: plan of investigations, specific medical treatment, non-medical aspects of management like surgical or psychosocial and major decisions regarding admission (e.g. ward vs ICU, isolation, discharge).

Pathophysiology including: etiology of disease, mechanisms or complications thereof, pathological findings, drug toxicity and complications of therapy.

Others can include: prevention of disease, public health issues, medical ethics, communication skills (like breaking bad news, giving consent).

All SCHS licensing examinations shall focus on important knowledge and competencies; while trivia, medical history and local protocols (unless widely accepted and relevant to KSA) should not be included. This becomes particularly true of licensing examinations, for example asking obscure anatomy serves no purpose, and presenting cases which will never be met in normal clinical practice does not inform about a candidate's readiness to practice.

Doctors taking the licensing exam most often have specialist diplomas, American Boards, UK fellowships and other postgraduate degrees. The exam itself is serving as a screen and as such it is not a recertification of the original specialist qualification, but a reasonable attempt to identify those practitioners who may not meet the minimum competencies expected for the safe treatment of the Saudi public. Therefore, the exam shall focus on knowledge and competencies central to that particular specialty.

Exams for postgraduate specialist qualifications represent a different type of exam. In order that they can be appropriate they must fulfill a series of objectives which the specialty states up front for all those in training. This difference must be appreciated by the item writer so that, if challenged, they can explain why the item informs the desired level of competence and experience at a licensing level.

When constructing items, there shall not be a slavish attempt to reproduce knowledge from the pages of a textbook as this will only serve to prove that the candidate has read that page. Instead the knowledge sought shall reflect what a specialist / doctor would do in normal every day practice, *supported* by the evidence in the textbook. However, it is worth remembering that most textbooks are out of date before they are even published in certain areas. The SMEs (subject matter experts) who create the items must use their own experience, which they can check against the stated texts, or published Evidence Based Medicine.

---

[5] The Test Blueprint has been devised by SCHS and will evolve with time to ensure satisfactory coverage for all exams

IMAGES, DIAGRAMS AND CARTOONS

The use of multimedia in a computer based examination is very useful as it becomes the nearest thing to a test of clinical competence that one can achieve without using a patient, simulated or otherwise, and an actual clinical situation.

However, there are some absolute necessities about the use of images; otherwise, their use is a distraction rather than an enhancement.

SCHS exams administered by Prometric are computer based and therefore the screen size and resolutions are issues for the display of images. **Q33** is an illustration of an inadequate image enhanced question. The resolution of the skull image is insufficient although the arrow is clearly visible. It can also be noted that the image was not even necessary as the skull foramen is correctly identified by the text under the image and then in the question line a form of colloquial English is used (*courses*), shown in red in the item box.

Q34 (Figure 2) shows the same question after editing. The test that it represents is still far from perfect but a considerable improvement on **Q33**. Note that because there are three structures which pass through the foramen ovale, and the options only include one, it is more appropriate to include the words: *which of the following structures* in the question line.

Figure 3 shows another screen shot of an image enhanced MCQ, **Q35**; the image is 920 pixels or 52 KB. This is quite sufficient for good computer viewing and is the minimum requirement. Other variants of image enhanced items are shown in Figure 4 (**Q36**) and Figure 5 (**Q37**).

It should not be forgotten that the use of a computer based assessment facilitates the use of coloured images, and this can, and should be, made maximum use of.

Figure1, Q33
An MCQ with an image. This illustrates several errors in construction, not least because the image is not even necessary. Note also the use of language in red.
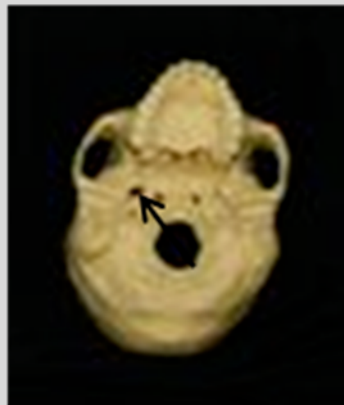
Refer to the accompanying picture of a skull with the foramen ovale indicated.

What nerve courses through this foramen?

A. Occulomotor
B. Maxillary
C. Mandibular
D. Abducens

Which of the following structures passes through the foramen?

A. Facial nerve
B. Ophthalmic vein
C. Meningeal artery
D. Mandibular nerve
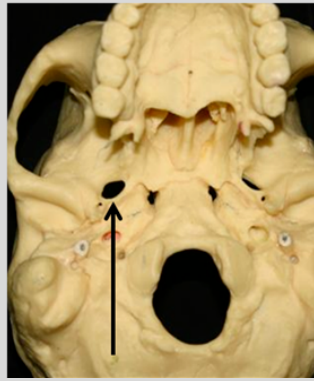
The base of the skull with a foramen indicated, through which passes an important structure.

Figure2, (Q34). Q33 has been edited and a more suitable image of a skull base used. Images are easy to create in one's own medical school and Web based images should be avoided, for which permission would almost always be needed to use to escape involving copyright issues.
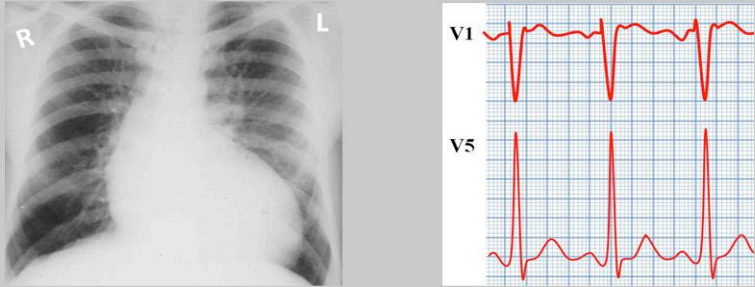
The patient complains of increasing tiredness and when his doctor takes his blood pressure finds it to be:

First reading   180/110 mmHg
Second reading   170/108 mmHg
Third reading   190/100 mmHg

What does the X-ray show?

A. Enlarged right atrium
B. Complete heart block
C. Left heart failure
D. Left ventricular hypertrophy

The patient's X-ray

Figure 3, (Q35). A screen shot of an MCQ with an image of suitable resolution. The size of the image and text on the screen will be approximately four times what is seen on this page.

Figure 4, (Q36). The two images above have been used to enhance an MCQ. The images could have been put one above the other or as here, alongside one another.

Figure 5, (Q37). An image enhanced MCQ used to illustrate the use of colour with good resolution.

REFERENCE SOURCE AND ADDED INFORMATION

The subject matter experts (SMEs) who create items must use their own experience, supported by evidence that is in internationally recognised textbooks. These references shall satisfy the following criteria:

1. Be from the main textbook sources in the field.
2. Book references shall be exclusively from international textbooks.
3. Only the latest published edition shall be used.
4. The number of textbook resources shall be limited to five, exceptionally may be increased to seven (SCHS will make recommendations from time to time).
5. Journal articles may not be used as sources for SCHS questions.

TABULATED DATA

In order to make an MCQ more easily readable with retention of the data, the use of tabulated data is highly recommended, including the presentation of the vital signs; see **Q38, Q39** and **Q40** as examples.

Q38
The use of tabulated data illustrated with the inclusion of arterial blood gases in an MCQ.

A 60 year-old man was brought to the Emergency Room by his wife with difficulty getting his breath for the last week, coughing up yellow-green sputum and forced to sit in a chair at night.

The doctor arranged an arterial blood gas (ABG) investigation which gave the following results:

Test	Results	normal values
pO_2	10	11-13 kPa
pCO_2	8.5	4.7-5.9 kPa
pH	7.25	
HCO_3	15	21-28 mmol/L

What is the correct interpretation of the ABGs?

    A. Respiratory alkalosis
    B. Compensated metabolic acidosis
    C. Decompensated metabolic acidosis
    D. Acute respiratory acidosis

Q39.
Another example of tabulated data, presented to illustrate the clarity that comes with this form of presentation.

A 42 year-old woman presented to her family doctor with tiredness and feeling light headed. The doctor arranged for some routine tests, below are the results:

Test	Result	Normal Range
Hb	80.0	120-160 g/L
MCV	82	86-98 fL
MCHC	240	320-360 g/L
MCH	24	28-33 pg
TIBC	43	45-66 mmol/L

A blood film was reported to show poikilocytosis and anisocytosis.

What type of problem does this patient have?

  A. Anaemia of chronic disease
  B. Thalassaemia trait
  C. Vit B$_{12}$ deficiency
  D. Iron deficiency

SCHS exams will always use SI units and provide normal values similar to the Medical Council of Canada, see Appendix **B**.

Or using the following URL:

http://www.mcc.ca/objectives_online/objectives.pl?lang=english&loc=values

Or, http://www.flash-med.com/LabNormal.asp

Q40.
Additional data such as results of special serological tests, imaging reports or those not fitting the presentation of lab data shall be given in their own paragraph as shown in blue here.

A 38 year-old woman complained of pain and stiffness of her left knee for three months. She felt tired and her sleep was disturbed by pain. On examination her knee was slightly swollen, felt warmer than the right and movement was restricted. Laboratory tests showed:

Test	Results	Normal ranges
Hb	100	120-160 g/L
MCH	30	28 – 33 pg
MCV	90	86-98 fL
TIBC	42	45-66 mmol/L
S. iron	12	11-29 mmol/l

The patient was IgM rheumatoid seronegative.

What was the most likely diagnosis?

  A. Palindromic rheumatism
  B. Osteoarthritis
  C. SLE
  D. Crystal arthritis

ITEM WRITING FLAWS

Although much has been written about item writing flaws (IWFs), they remain all too frequent at all levels of examination, from classroom tests to licensing examinations. SCHS exams must be screened to eliminate these flaws in the belief that examinations with a significant proportion of flaws in a single exam will produce invalid results (Tarrant M et al, 2009).

Already a number of IWFs have been discussed in the preceding sections as their presence causes items to lose clarity and the questions asked to become ambiguous. Appendix **C** provides nineteen of the classic IWFs frequently quoted in the literature. Eight of the IWFs from Appendix **C** will be discussed below with examples.

Avoid Gratuitous information

The most obvious example of gratuitous information in an examination for overseas doctors is the use of hospital, place and person names as these will often mean little to the candidate, see **Q41**.

Q41.
Care must be taken to ensure that every fact stated will be used to answer the question in the question line.

> Abdulla Mohammed, a successful 36 year-old actuary working for Price Cooper Waterhouse in Abha, went to the Arabic Nation's Hospital near the airport. He has been experiencing palpitations for the last two weeks after returning from a trip to New York, via London and Riyadh. The female Emergency Room doctor arranged a series of routine tests, including a chest X-ray, ECG, serum troponin levels and background Haematology. None of these tests came back abnormal.
>
> What did the Emergency Room doctor initially suspect that the patient was suffering from?
>
>   A. Silent infarction
>   B. Acute anxiety
>   C. Vitamin disorder
>   D. Overwork

The absurdity of most of the information in **Q41** is obvious but is included to reinforce the point being made. Do not use names, do not enumerate each individual investigation ordered if they all turn out normal, use summaries. **Q41** has a wpi = 95, while the actual number required is half, see **Q42**, wpi = 50.

Q43 is another example of an item where <u>all</u> the information in the vignette is completely unnecessary. This error occurs more frequently than we think and may be very similar to logical clues and word repeats in the stem and correct option, see **Qs 43-44**. This allows a candidate, who is either confused or does not have the knowledge base, to guess with a 50-50 chance of being correct.

Q42.
Q41 has been edited to exclude the irrelevant information. Now only the information necessary to answer the question remains.

> A 36 year-old male has just returned from a long and stressful journey, while experiencing palpitations for the last two weeks. The cardiac screening tests ordered in the Emergency Room were all normal.
>
> What would be the most likely diagnosis?
>
> A. Silent infarction
> B. Acute anxiety
> C. Vitamin disorder
> D. Overwork

Q43.
The details given in the vignette are unnecessary as the diagnosis has been given away in the question line.

> A 38 year-old woman complained of pain and stiffness of her left knee for three months. She felt tired and her sleep was disturbed by pain. On examination her knee was slightly swollen, felt warmer than the right and movement was restricted. Laboratory tests showed:
>
> | Test | Results | Normal ranges |
> |------|---------|---------------|
> | Hb | 100 | 120-160 g/L |
> | MCH | 30 | 28 – 33 pg |
> | MCV | 90 | 86-98 fL |
> | TIBC | 42 | 45-66 mmol/L |
> | S. iron | 12 | 11-29 mmol/l |
>
> What investigation would be most appropriate for this case of Rheumatoid Arthritis?
>
> A. Joint aspiration
> B. IgM serology
> C. X-ray of knee
> D. Plasma Electrophoresis

Q44.
There is a logical clue inadvertently left in the vignette and also two of the options, one of which is the correct key.

> A 22 year-old woman experienced twenty minutes of facial pain and left sided arm weakness, followed by a throbbing right sided headache. After four hours, she felt normal again. The next day, the doctor could find no abnormal findings on examination.
>
> What was the most likely diagnosis?
>
> A. TIA
> B. Tension headache
> C. SAH
> D. Migraine

SMEs should be mindful of these types of flaw and how to avoid them through their own editing. **Q45** is an example of how IWFs in **Q44** can be eliminated. Clearly the area being tested is important and can still be tested.

Q45.
Q44 has been modified to place all the options on the same continuum with headache. Although this does not entirely exclude the use of logic, word repeats are excluded with four options to choose from.

> A 22 year-old woman experienced twenty minutes of facial pain and weakness of the left arm, followed by a throbbing right sided headache. After four hours, she felt normal again. The next day, the doctor could find no abnormal findings on examination.
>
> What was the most likely diagnosis?
>
>   A. Cluster headache
>   B. Tension headache
>   C. Post-trauma headache
>   D. Migraine

Q46.
An example of an MCQ which can be answered by just using logic.

> A previous healthy 36 year-old man presented with a two day febrile illness and tiredness starting ten days after he had returned from India. On examination he had scleral icterus and mild right hypochondrial tenderness.
>
> What is the most likely diagnosis?
>
>   A. Chronic HCV
>   B. Gilbert's syndrome
>   C. Acute viral hepatitis
>   D. Chronic HBV

Q46 suggests the man has an infection, which started after his return from India. Logically this is an acute problem. Therefore, options **A**, **B** and **D are** excluded. Although one might argue that by excluding **A**, **B** and **D**, implies some level of knowledge, the level is what one would expect of an undergraduate medical student not a doctor applying for a license to practice independently. More information is required of such a doctor.

Q47 shows quite a popular way of testing for memorized facts at the level of first and second medical school years. This is not suitable for licensing examinations of whatever type. The correct key is almost always written first and then the three distractors fitted around it by adding one or two different facts, but all the time increasing the number count for those included in the correct key. Please note five options have been used to illustrate the point being made.

Note the number of times the items are repeated in the option list.

- Ruptured tibiofibular (sic ligament):   3 times
- Medical ligament (deltoid):   4 times
- Fracture of distal fibula: 3 times

Therefore in option **C**, the same injuries occur ten times, whereas the next greatest is nine which leaves option **C** as the convergent option and correct key.

Q47.
A further example of an IWF, with a very obvious convergence clue.

> What is the associated bone and ligament damage with abduction injuries of the ankle joint?
>
> A. Ruptured tibiofibular ligament (3) and fractured medial (2) and posterior malleolus (1)
> B. Ruptured medial (3) and lateral ligaments (2) with a fracture of the distal fibula (3)
> C. Ruptured tibiofibular (3) and medial ligaments (4) with a fracture of the distal fibula (3)
> D. Medial malleolar fractures (2) with tibiofibular (3) and medial ligament rupture (3)
> E. Ruptured medial (4) and lateral ligaments (2) and fractured distal fibula (3).

Perhaps the best known IWF is the longest option being best, although true statistically, it is the correct key 50-60% of the time due to the fact that qualifying clauses are added to ensure that the option is truly correct, see **Q48**.

Q48.
The best known IWF, where the longest option is the correct key. This is true 50-60% of the time due to the extra qualifying information included with the correct keyed option.

> A 78 year-old man woke in the middle of the night gasping for breath due to a feeling of suffocation. After sitting upright on the side of his bed for about five minutes he felt more comfortable, and then placing three pillows under himself and lying in a semi reclining position, he soon fell asleep again. *He was relieved that the district nurse who calls regularly gave him this advice after it happened the first time.*
>
> What is the most likely cause of the patient's condition?
>
> A. Acute right sided heart failure
> B. Sleep apnoea and cerebral hypoxia
> C. Early changes of respiratory failure
> D. Paroxysmal nocturnal dyspnoea and associated pulmonary oedema

Note too, that in **Q48** there is redundancy shown in red italics. The addition of these extra facts have no bearing on the information needed to answer the question in the question line.

Q49 and **Q50** illustrate the ascending order of the options. SCHS does not require alphabetically listed options as with **Q49**.

Q49.
It is not considered necessary to place options in an alphabetical order as the computer will randomly deliver the options in at least four different orders (twelve are possible).

A 40 year-old investment banker has had three months of epigastric pain, which even woke him at night. He is a heavy smoker and consumer of alcohol and admitted to much stress in work, but denied taking brufen or any similar non-prescription drug. His own doctor prescribed a proton pump inhibitor and referred him to the specialist clinic where duodenal ulcer disease is confirmed, but the rapid urease test on an antral biopsy was reported negative.

What is the most likely cause of the ulcer?

    A. Alcohol
    B. *H. pylori*
    C. Stress
    D. Tobacco

Q50.
Items using range data shall always be presented in an ascending or descending order with no overlap.

A 42 year-old man presented after a short flu like illness with mild jaundice and increasing tiredness. He denied use of intravenous drugs and has no history of a recent blood transfusion. Lab tests showed:

Test	Results	Normal values
Hb	74	140-180 g/L
TIBC	46	45-66 mmol/L
S. iron	9	11-29 mmol/l
S. Ferritin	135	5.8–120 nmol/L
ALT	0.62	0.0-0.58 mkat/L
Bilirubin	34	5.1-17.0 mmol/L

It was decided to commence Pegylated α-2b interferon and Ribavirin.

Without this treatment what is the approximate risk of developing chronic liver disease?

    A. 10 – 20%
    B. 30 – 45%
    C. 50 – 65%
    D. 70 – 85%

Q51, is a format to avoid if possible, seldom in biological sciences does anything conform to an exact measurement or proportion. Ranges are more realistic. And **Q51** could be edited accordingly. **Q52** is acceptable, although the measurements are not stated in ranges, because the candidate will be performing a calculation using a formula, which will always yield exact numbers.

Q51.
Although correct in ascending order, the exact nature of measurements in the options make each unlikely and therefore false.

At the level of the C-1 vertebra, what percentage of the spinal canal is occupied by the spinal cord?

    A. 12.5%
    B. 25%
    C. 33%
    D. 50%

Q52.
Despite exact numbers again, these might be considered more plausible and acceptable.

A doctor notes that the right atrial pressure is 12 mmHg (13593 $Kg.m^3$) *** in an adult male.

How many cms above the sternal angle would the "**a** wave" be seen with the patient sitting up at 45°?

    A. 2.5
    B. 5.6
    C. 7.4
    D. 11.2

***, Specific Gravity = weight / volume

Q53.
Illustration of absolute terms that are always rejected as being the correct answers.

A 38 year-old woman with diabetes mellitus (DM) for the last ten years and treated with oral antiglycaemic agents, asked the specialist what her chances were of developing serious kidney disease.

What should the doctor answer?

    A. Only type 1 DM develop kidney disease
    B. Insulin treatment always increases the risk
    C. Poor control never reduces the risk
    D. Develops irrespective of type of treatment

Q53, illustrates absolutes qualifying three options (**A**, **B** and **C**). This should be avoided as absolutes are seldom true in biological sciences. **Q53** also illustrates another flaw that will cause confusion, the use of double negatives in the same option, **C**; *poor control never reduces the risk*. This should read: *Control does not affect risk*: one fewer negative in the statement. Similar misuse of English is the use of vague terms which can mean different values to different people. Below are the opinions of a Faculty office asked what they thought about vague terms (their estimates were averaged):

- Never – considered to occur at least 5%-10% of the time, a variation of 100%
- Rarely occurred – thought to occur about 20% of the time
- Always took place – meant at least 80% of the time
- Often happened – meant whatever it was that happened, it did so 60-70% of the time
- Sometimes it happened – meant it took place 40% of the time.
- Invariably took place – meant about 50% of the time.

If one must, the absolutes can be removed as they seldom change the sense of the item, see **Q54**. However, quantification is preferable and more acceptable.

Q54.
Illustration of removal of absolutes and little else changed, seemingly not to affect the value of the options.

A 38 year-old woman with type II diabetes mellitus (DM) for the last ten years and treated with oral antiglycaemic agents asked the specialist what her chances were of developing serious kidney disease.

What should the doctor answer?

  A. Type 1 DM develops kidney disease
  B. Insulin treatment will increases the risk
  C. Glycemic control not a risk factor
  D. Unaffected by type of treatment

Vague terms and absolute qualifying statements in an option lead candidates away from that option and thereby to the correct key.

Q55 is an example of options that are far too complex and would waste many valuable seconds to read through to fix in a candidate's memory.

Q55.
Confusing, complex and takes too long to comprehend. An example of option lengths being too long

A 20 year-old motorcyclist crashed into a stationary van and was immediately rushed to the nearest Hospital. He sustained a burst fracture of the C5, was paralysed below C6 but has Grade 2/5 flexor hallucis power on the right side and was able to contract his anal sphincter.

Which of the following is most appropriate in this case?

  A. Methylprednisolone bolus 30 mg/kg within 6-8 hours of injury followed by a continuous infusion of 5.4 mg/kg for 23 hours

  B. Steroid therapy is not indicated because of an incomplete lesion

  C. Arrange an immediate MRI as indication and dosage of steroid would depend upon MRI findings and presence/absence of concomitant head injury

  D. Immediate spinal stabilization will prevent further neurological deterioration and must be arranged as emergency

Q56 and **Q57** are examples of two MCQ forms that SCHS will not use. In the case of *none of the above* (NOTA), if any of the three options **A**-**C** are remotely true then **D** is not the correct key and is eliminated, thereby making the item easier. Similarly for **Q57**, all of the above (AOTA), if just one is not true then option **D** will not be true and the item becomes correspondingly easier.

Q56.
An MCQ with the final
option stating *none of
the above* (NOTA).
Candidates usually
eliminate the NOTA
option as a filler.

A patient complained of facial flushing, made worse by hot drinks and more recently noticed small papules and pustules on his forehead and nose. The patient wore reading glasses.

What is the treatment to be recommended?

    A. Prednisolone 10 mg daily
    B. Isotretinoid 2.0 mg/kg/day
    C. Oral tetracycline 150 mg daily
    D. None of the above

Q57.
An MCQ with the final
option stating *all of the
above* (AOTA).
Candidates usually
eliminate the AOTA
option as a filler.

A male patient has just been diagnosed with polycystic disease of the kidneys. After hearing about the familial nature of the disease he asked which family members may have the disease. The patient has two brothers and two sisters and their mother was also known to have the disease.

What should the doctor tell him?

    A. His brother
    B. One of his sisters
    C. The two sisters
    D. All of the above

APPENDIX A

THE TEST BLUEPRINT

A Test Blueprint is a guide that is used to create an examination and facilitate wide sampling from the published list of the competencies and topics. Test Blueprints come in a variety of different forms. Illustrated in Table 1 is the information provided by SCHS to the item writers for **Paediatric Surgery**. The goal of a Test Blueprint (TB) is to ensure as wide and representative sample as possible is taken of a candidate's knowledge in a particular area or subject.

Table 1 is an example of a Test Blueprint for paediatric surgery, 200 items requested. Under each of the headings: <u>Diagnosis</u> include: clinical reasoning, interpretation of symptoms and/or signs to reach the most likely diagnosis, proper interpretation of investigations. <u>Management</u> include: plan of investigations, specific medical treatment, non-medical aspects of management like surgical or psychosocial and major decisions regarding admission (e.g. ward vs ICU, isolation, discharge). <u>Pathophysiology</u> include: etiology of disease mechanisms or its complications, pathological findings, drug toxicity and complications of therapy. <u>Others</u> can include: prevention of disease, public health issues, medical ethics, communication skills (like breaking bad news, giving consent). It can be noted that percentages are not given as these would anyway need to be converted into real numbers, as shown below.

Table 1. Example of a Test Blueprint for paediatric surgery, 200 items requested.

MAIN HEADINGS	SUBHEADINGS				
	TOTALS	Diagnosis	Management	Pathophysiology	Others
General	24	8	8	6	2
Trauma	20	6	6	6	2
Oncology	30	8	8	8	6
Transplantation	6	2	2	2	-
Head & neck	10	4	4	2	-
Thorax	30	10	10	8	2
Abdomen	50	14	12	12	12
Genitourinary	20	6	6	6	2
Special	10	4	4	2	-
TOTALS	200	62	60	52	26

The four headings used under Subheadings, Table 1, are summaries for what is included and the detailed areas covered are given in the table legend. SCHS also supplies more detail under Topic Areas and Appendix A is an example for Internal Medicine.

Several other parametres can be used to ensure that the test coverage is appropriate, such as the proportion of K1 and K2 items[6]. For most purposes this can be an across the board criteria for all SCHS exams and does not have to be applied differentially under the various main and subheadings. The policy for the exam will clearly state the requirement.

In order to achieve testing of higher cognition the NBME policy has been to use vignettes, as this subject will be addressed in its own section of this manual it will only be briefly referred to here. However, already the difference between *setting* a context and *giving* a context shall be made with two examples. Example **QA1** is an example of a K1, recall item, with the context being *set*, vasospasm associated with subarachnoid haemorrhage.

QA1.
This question illustrates the context being set, vasospasm following a subarachnoid haemorrhage, while at the same time only testing at the lowest cognitive level.

> A 23 year-old has sudden onset of severe headache and loss of consciousness. On examination there was a right sided hemiparesis secondary to a subarachnoid haemorrhage.
>
> Which of the following is an EBM Class I management option to prevent vasospasm in this condition?
>
> > A. Hypervolaemia
> > B.Heparin
> > C.Statins
> > D. Vasodilators

QA2 shows how the same question has been presented but this time it will take clinical reasoning to find the best answer. The context is given with clinical data and must be used to answer the question line.

QA2.
This is an example of an item being *given* a clinical context using patient data. Importantly the clinical data must be used to answer the question asked.

> A 23 year-old has a sudden onset of severe headache and loss of consciousness. On examination there was a right sided hemiparesis. A CT scan confirmed an intracranial haemorrhage and an angiogram a narrowed MCA.
>
> What immediate treatment will increase the rate of survival?
>
> > A. Hypervolemic
> > B.Heparin
> > C.Statins
> > D. Vasodilators

---

[6] Bloom described six levels in the taxonomy of the cognitive domain, recall, comprehension, application analysis, synthesis and evaluation. However, a panel of experts have difficulty in reaching consensus about specific items at this level of specificity and therefore most Q-bank classifications use only recall and reasoning, or K1 and K2.

Examples of topic areas provided by SCHS for Internal Medicine.

1. **Cardiovascular Disease**

Ischaemic heart disease (IHD)
Valvular heart disease
Cardiomyopathies
Heart failure (LVF, CHF)
Hypertension
Pericardial disease

2. **Pulmonary Diseases**

Chronic obstructive pulmonary disease (COPD)
Bronchial asthma
Interstitial lung disease and pulmonary fibrosis
Bronchiectasis
Ventilation management and oxygen therapy
Adult respiratory distress syndrome
Pleural effusion.

3. **Rheumatology**

Crystal induced arthropathy
Acute monoarthritis
Rheumatoid arthritis
Systemic lupus erythematosus
Vasculitis
Osteoarthritis
Metabolic bone disease
Back pain
Miscellaneous Rheumatology

4. **Oncology**

Lung cancer
Lymphoma
Breast cancer
Hepatocellular carcinoma and G.I. malignancies
Prostate cancer
CNS tumors
Metastatic cancer with unknown primary
Paraneoplastic syndrome

5. **Nephrology**

Acute renal failure
Chronic kidney disease
Glomerulonephritis

Water and electrolyte disturbance
Acid-base disturbance

6. **Infectious Diseases**

Community-acquired pneumonia (CAP)
Tuberculosis (pulmonary and extra-pulmonary)
Brucellosis
Malaria
HIV and AIDS
Urinary tract infection
Skin and soft-tissue infections
Sepsis and septic shock
Septic arthritis
Infections in Immuno-compromised patients
 (post-transplant, neutropenic patients)
Health-care associated infections and infection
 control
CNS infections
Specific organism infections like Staph aureus,
group A Strep,
Upper-respiratory infections
Acute viral infection like influenza, infectious
mononucleosis
Active enteric infections
Acute and chronic viral hepatitis
Infective endocarditis

7. **Hematology**

Anemias including hemolytic anemias
Haemoglobinopathies (Sickle-cell disease,
 Thalassemia)
Coagulopathies venous thrombo-embolic
 disease (VTE)
Transfusion medicine
Myeloproliferative and myelodysplastic disorders
Hypersplenism and pancytopenia

8. **Gastroenterology**

Peptic ulcer disease
Gastrointestinal bleeding
Esophageal diseases
Chronic inflammatory bowel disease
Biliary tract diseases
Malabsorption syndromes / celiac disease

9. Endocrine and metabolism

Diabetes mellitus
Lipid disorders
Thyroid gland disorders
Metabolic bone disease
Obesity
Miscellaneous endocrine disorders

10. Emergency Medicine

Acute coronary syndrome
Pulmonary edema
Cardiogenic shock
Massive pulmonary embolism
Respiratory failure
Status asthmaticus
Status epilepticus
Pneumothorax
Gastrointestinal bleeding
Acute liver failure
Diabetic ketoacidosis
Hyperosmolar non-ketotic state
Adrenal crisis
Acute monoarthritis
Common drug overdose
Septic shock
Acute bacterial meningitis
Severe electrolyte disturbance (K, Na, Ca)

11. Medical Therapeutics

Mechanism of action, indication, and main side effects of commonly used medications like NSAIDS, antibiotics, and hypoglycemic, insulin anti-epileptics, drugs used for hyperlipidemia, cardiac medication, diuretics, etc.

12. Neurology

Cerebrovascular accident (stroke, cerebral hemorrhage)
Seizures and epilepsy
Multiple sclerosis
Parkinson's Disease
Peripheral neuropathy
Guillain-Barre syndrome
Rehabilitation

13. Psychiatry

Common diseases that the internist needs to know.

14. Dermatology

Commonly seen skin rashes particularly those related to medical conditions.

APPENDIX B

(SI Units used for SCHS):

Blood Test	**SI Units**
1,25-dihydroxyvitamin D [1,25(OH)$_2$D]	40-160 pmol/L
11-Deoxycortisol (compound S)	<30 nmol/L
17 Hydroxyprogesterone Men	0.2-9.0 nmol/L
17 Hydroxyprogesterone Women Follicular phase	0.6-3 nmol/L
17 Hydroxyprogesterone Women Luteal phase	1.5-10.6 nmol/L
25-hydroxyvitamin D [25(OH)D]	20-200 nmol/L
Acetoacetate	<0.3 mmol/L
Adrenocorticotropin (ACTH), 8 A.M.	7-27-27 pmol/L
Alanine Aminotransferase (ALT, SGPT)	0-0.58 mkat/L
Albumin	35-55 g/L
Aldolase	<135 nkat/L
Aldosterone	14-53 nmol/d
Aldosterone, 8 A.M., **	<220 pmol/L
Alpha fetoprotein (adult)	<30 mg/L
Alpha$_1$ antitrypsin	0.8-2.1 g/L
Ammonia, as NH$_3$	0-47 mmol/L
Amylase	0.8-3.2 mkat/L
Androstenedione Men	3.0-5.0 nmol/L
Androstenedione Women	3.5-7.0 nmol/L
Angiotensin II, 8 A.M.	10-30 nmol/L
Angiotensin-converting enzyme (ACE)	<670 nkat/L
Antinuclear antibody agglutination	<1:80 titer
Arginine vasopressin (AVP), random fluid intake	1.4-5.6 pmol/L
Arterial blood gas [HCO$_3^-$]	21-28 mmol/L
Arterial blood gas P$_{CO2}$	4.7-5.9 kPa
Arterial blood gas pH	7.38-7.44
Arterial blood gas P$_{O2}$	11-13 kPa
Ascorbic acid (Vitamin C)	23-57 mmol/L
Aspartate Aminotransferase(AST, SGOT)	0-0.58 mkat/L
AVP (Arginine vasopressin), random fluid intake	1.4-5.6 pmol/L
Bands	0-0.04
BASO (basophils)	0-0.02
Basophils (BASO)	0-0.02
Beta -Hydroxybutyrate	<300 mmol/L
Bilirubin, Direct	1.7-5.1 mmol/L
Bilirubin, Indirect	3.4-12 mmol/L

Bilirubin, total	5.1-17 mmol/L
Bleeding time (Simplate)	<7 min
Calcitonin Men	<13.8 ng/L
Calcitonin Women	<6.4 ng/L
Calcium	2.25-2.62 mmol/L
Calcium, ionized	1.1-1.4 mmol/L
Carbon dioxide content	23-28 mmol/L
Carbon dioxide tension (P_{CO_2})	4.7-5.9 kPa
Carboxyhemoglobin non-smoker	0-0.023
Carboxyhemoglobin smoker	0.021-0.042
Carotene	1.4-5.6 mmol/L
Carotenoids	0.9-5.6 mmol/L
Ceruloplasmin	270-370 mg/L
Chloride	98-106 mmol/L
Cholesterol, HDL (Acceptable)	>1.03 mmol/L
Cholesterol, HDL (Borderline depressed)	0.9-1.0 mmol/L
Cholesterol, HDL (Depressed)	<0.9 mmol/L
Cholesterol, LDL (Acceptable)	<3.36 mmol/L
Cholesterol, LDL (Borderline elevated)	3.36-4.11 mmol/L
Cholesterol, LDL (Elevated)	4.14 mmol/L
Cholesterol, Total (Acceptable)	3.88-5.15 mmol/L
Cholesterol, Total (Borderline elevated)	5.17-6.18 mmol/L
Cholesterol, Total (Elevated)	>6.2 mmol/L
Chorionic gonadotropin, Beta-subunit (B-hCG), men and non-pregnant women	<3 IU/L
Copper	11-24 mmol/L
Cortisol A.M.	140-690 nmol/L
Cortisol P.M.	80-330 nmol/L
Cortisol Urinary Free	25-140 nmol/d
Creatine kinase-MB	<0.03 of total
Creatinine	<133 mmol/L
Creatinine kinase (CK) Females	<3.17 mkat/L
Creatinine kinase (CK) Males	<3.92 mkat/L
Dehydroepiandrosterone (DHEA)	7-31 nmol/L
Delta-aminolevulinic acid	<1.5 mmol/L
DHEA (Dehydroepiandrosterone)	7-31 nmol/L
DHEA sulfate	1.3-6.8 mmol/L
EOS (eosinophils)	0-0.07
Eosinophils (EOS)	0-0.07
Erythrocyte count	$4.15\text{-}4.90 \times 10^{12}$/L
Erythropoietin	5-36 U/L
Estradiol Men	<180 pmol/L

Estradiol Women (higher at ovulation)	70-220 pmol/L
Ethanol	<0.005%
Euglobulin lysis time	>2 h
Fatty acids, free (non-esterified)	180 mg/L
Ferritin Men	15-400 mg/L
Ferritin Women	10-200 mg/L
Fibrinogen	1.5-4 g/L
Folate	>4.3 nmol/L
Folate, red cell	>95 nmol/L
Folic acid	340-1020 nmol/L cells
Folic acid	7-36 nmol/L cells
Follicle-stimulating hormone (FSH) Men	0.9-15 IU/L
Follicle-stimulating hormone (FSH) Postmenopausal Women	34-96 IU/L
Follicle-stimulating hormone (FSH) Women Ovulatory surge	2.3-21 IU/L
Follicle-stimulating hormone (FSH)✻	1.4-9.6 IU/L
FSH (Follicle-stimulating hormone)	0.9-15 IU/L
Gastrin	40-180 ng/L
Glucagon	50-100 ng/L
Glucose fasting (Diabetes mellitus)	>7.8 mmol/L
Glucose fasting (Normal)	4.2-6.4 mmol/L
Glucose--2 hour post prandial (Impaired)	7.8-11.1 mmol/L
Glucose--2 hour post prandial (Normal)	<7.8 mmol/L
Glucose--2 hour postprandial (DM)	>11.1 mmol/L
Growth hormone, after 100 g oral glucose	<2 mg/L
Ham's test (acid serum)	Negative
Haptoglobin (serum)	0.5-2.2 g/L
HbF (Hemoglobin, fetal)	<0.02
HCT Females (hematocrit)	0.37-0.48
HCT Males (Hematocrit)	0.42-0.52
Hematocrit Females (HCT)	0.37-0.48
Hematocrit Males (HCT)	0.42-0.52
Hemoglobin A_{1c}	0.038-0.064
Hemoglobin A_2 (HbA_2)	0.015-0.035
Hemoglobin Female	120-160 g/L
Hemoglobin Male	140-180 g/L
Hemoglobin Whole blood Females	7.4-9.9 mmol/L
Hemoglobin Whole blood Males	8.1-11.2 mmol/L
Hemoglobin, fetal (HbF)	<0.02
Insulin, fasting	43-186 pmol/L
Iron	11-29 mmol/L
Iron Saturation	0.2-0.45

Iron-binding capacity	45-66 mmol/L
Lactate (Lactic acid)	0.67-1.8 mmol/L
Lactate dehydrogenase (LDH)	1.7-3.2 mkat/L
Lactate dehydrogenase isoenzymes Fraction 1*	0.14-0.25
Lactate dehydrogenase isoenzymes Fraction 2*	0.29-0.39
Lactate dehydrogenase isoenzymes Fraction 3*	0.20-0.25
Lactate dehydrogenase isoenzymes Fraction 4*	0.08-0.16
Lactate dehydrogenase isoenzymes Fraction 5*	0.06-0.16
LAP (leukocyte alkaline phosphatase)	0.2-1.6 mkat/L
LDH (Lactate dehydrogenase)	1.7-3.2 mkat/L
Lead	<1.9 mmol/L
Leukocyte Alkaline phosphatase (LAP)	0.2-1.6 mkat/L
Leukocyte Count (WBC, white blood cell count)	$4.3\text{-}10.8 \times 10^9$/L
Lipase	0-2.66 mkat/L
Lipoprotein (a)	0-300 mg/L
Lutenizing hormon Men	1.3-13 IU/L
Lutenizing hormone Women Mature, premenopausal, except at ovulation	0.8-26 IU/L
Lutenizing hormone Women Ovulatory surge	25-57 IU/L
Lutenizing hormone Women Postmenopausal	40-104 IU/L
Lymphocytes (Lymphs)	0.16-0.45
Lymphs (lymphocyes)	0.16-0.45
Magnesium	0.8-1.2 mmol/L
MCH	28-33 pg/cell
MCHC	320-360 g/L
MCV	86-98 fl
Mean corpuscular hemoglobin (MCH)	28-33 pg/cell
Mean corpuscular hemoglobin concentration	320-360 g/L
Mean corpuscular volume (MCV)	86-98 fl
Mono (monocytes)	0.04-0.10
Monocytes (Mono)	0.04-0.10
Myoglobin	<55 mg/L
Neutrophils (Poly, PMN)	0.45-0.74
Osmolality	285-295 mmol/kg
Oxygen percent saturation (sea level)	0.97 mol/mol
Oxygen tension (P_{O2})	11-13 kPa
Parathyroid hormone	10-60 ng/L
Parathyroid hormone-related protein	<1.3 pmol/L
pH	7.38-7.44
Phosphatase, acid	0.90 nkat/L
Phosphatase, alkaline	0.5-2.0 nkat/L
Phosphorus, inorganic	1.0-1.4 mmol/L
Platelets	$130\text{-}400 \times 10^9$/L

PMN (neutrophils, Poly)	0.45-0.74
Poly (neutrophils, PMN)	0.45-0.74
Potassium	3.5-5.0 mmol/L
Progesterone Men, prepubertal, preovulatory and postmenopausal females	<6 nmol/L
Progesterone Women, luteal, peak	6-60 nmol/L
Prolactin	2-15 mg/L
Prostate specific antigen >40 years	0.0-4.0 mg/L
Prostate specific antigen Male: <40 years	0.0-2.0 mg/L
Prostate-specific antigen (PSA)Female	<0.5 mg/L
Protein, total	55-80 g/L
Protoporphyrin, free erythrocyte (FEP)	0.28-0.64 mmol/L RBCs
PSA, free, in males 45-75 years, ***	>0.25 with BPH[§]
Pyruvate	60-170 mmol/L
RDW (Red cell distribution width)	0.13-0.15
Red cell distribution width (RDW)	0.13-0.15
Reverse T_3 (rT_3)	0.04-0.29 nmol/L
Rheumatoid factor	<40 kIU/L
rT3 (Reverse $T_{3)}$	0.04-0.29 nmol/L
Sodium	136-145 mmol/L
Sucrose hemolysis	Negative
$T_{3\ (Free)}$	3.5-6.5 pmol/L
T3 (Total)	0.9-2.8 nmol/L
Testosterone Men	10-35 nmol/L
Testosterone Women	<3.5 nmol/L
Thyroglobulin	0-60 mg/L
Thyroid stimulating hormone (TSH)	0.4-5.0 mU/L
Thyroxine (T_4)	64-154 nmol/L
Transferrin	2.3-3.9 g/L
Triglycerides	<1.8 mmol/L
Troponin I	0-0.4 mg/L
Troponin T	0-0.1 mg/L
TSH (Thyroid stimulating hormone)	0.4-5.0 mU/L
Urea nitrogen	3.6-7.1 mmol/L
Uric acid Men	150-480 mmol/L
Uric acid Women	90-360 mmol/L
Vitamin A	0.7-3.5 mmol/L
Vitamin B_1 (thiamine)	0-75 nmol/L
Vitamin B_{12}	148-443 pmol/L
Vitamin B_2 (riboflavin)	106-638 nmol/L
Vitamin B_6	20-121 nmol/L

Vitamin C (ascorbic acid)		23-57 mmol/L
Vitamin D$_3$,1,25-dihydroxy		60-108 pmol/L
Vitamin D3,25-hydroxy	Winter	34.9-105 nmol/L
Vitamin D3,25-hydroxy	Summer	37.4-200 nmol/L
Vitamin E		12-42 mmol/L
von Willebrand's antigen		60-150%
WBC (leukocyte count, white blood cell count)		4.3-10.8 $\times$ 10^9/L
Westergren Sed Rate Females < 50 y/o		0-20 mm/h
Westergren Sed Rate Females > 50 y/o		0-30 mm/h
Westergren Sed Rate Males < 50 y/o		0-15 mm/h
Westergren Sed Rate Males >50 y/o		0-20 mm/h
White blood cell count (leukocyte count, WBC)		4.3-10.8 $\times$ 10^9/L
Zinc		11.5-18.5 mmol/L

✱, Women Mature, premenopausal, except at ovulation

*, of total

**, (patient supine, 100 mmol/L Na and 60-100 mmol/L K intake)

***, with PSA values between 4 and 20 mg/mL

§BPH, benign prostatic hyperplasia

APPENDIX C

The advantages of using multiple choice question formats are well documented, what is not so well reported are the potential pitfalls and consequences of their use. Below is a list of item writing flaws (IWFs) followed by a brief explanation as a reminder of these consequences.

The research literature is divided about whether or how important some of the flaws are for the outcome of an MCQ examination.

Several of the IWFs will automatically be avoided if the in-house style (described in the *Style Guide*) is adhered to, such as #1, 2, 3, 4, 6 and 19.

A quick check list

1. All options should be grammatically consistent with the stem and should be parallel in style and form. Non-grammatically correct options provide cues to the students who easily eliminate distracters that do not flow grammatically with the stem.

2. Each MCQ should have a clear and focused question. Item writers <u>should not</u> construct MCQs with unfocused stems which then do not ask a clear question or state a clear problem in the sentence completion format.

3. Each MCQ should have the problem in the stem of the question, not in the options. And, the options should not be a series of true/false statements.

4. The basic format for MCQs is the single best answer. Therefore, ensure that questions have one, and only one, <u>best answer</u>.

5. Avoid gratuitous or unnecessary information in the stem or the options. If a vignette is provided with the MCQ, it should be required to answer the question.

6. Do not construct complex, or K-type MCQs. K-type MCQs have a range of correct responses and then require candidates to select from a number of possible combinations of these answers. Examinees can often guess the answer by eliminating one incorrect response and all options containing by selecting the responses which appear most frequently in all of the options.

7. Questions and all options should be written in clear, unambiguous and *simple* language[7]. Poorly worded or ambiguous questions can confuse even knowledgeable students and cause them to answer incorrectly.

8. Make all distracters plausible, as <u>plausible distracters are vital to high quality MCQs</u>.
    Students who do not know the material increase their chances of guessing the correct option by eliminating implausible distracters.

9. Do not repeat words in the stem and the correct option if it is possible to find appropriate and relevant replacements. Similar wording allows candidates to identify the correct option without knowing the material.

10. Avoid providing logical cues in the stem and the correct option that can help the examinees to identify the correct option without knowing the material. An example of a logical cue is asking students to select the most appropriate pharmaceutical intervention for a problem and only having one or two options which are actual pharmaceutical interventions.

11. Always seek ways to avoid convergence cues in options: that is where there are different combinations of multiple components to the answer. Question writers tend to use the correct answers more frequently across all options. While examinees identify the correct option in which all components appear most frequently.

12. All options should be similar in length and amount of detail. If one option is longer, includes more detailed information, or contains more complex language, students can usually assume that this is the correct answer (50%-60%) most of the time.

13. Arrange MCQ options in a logical, chronological, or numerical order. (SCHS assess for chronological and numerical, <u>but not alphabetical order</u>).

14. Options should be worded to avoid the use of absolute terms (e.g., never, always, only, all) as every student knows that there are often no absolute truths in most health science subjects and they can therefore eliminate these distracters. This also applies to size, time or amounts; use ranges that the correct answer will fall within.

---

[7] Simple language is without jargon, colloquialisms or local / regional specific expressions or use of words

15. Options should be worded to avoid the use of vague terms (e.g., frequently, occasionally, rarely, usually, commonly) as these terms lack precision and there is seldom agreement on the actual meaning of "often" or "frequently."

16. Do not use negatively phrased questions (e.g., not, except, incorrect, least) in the lead-in as they poorly assess examinees' knowledge. If the item assesses contraindications, the questions should be worded clearly to indicate that this is what is being assessed.

17. Do not use "all of the above" as the last option. Examinees can easily identify if this is the correct answer by simply knowing that at least two of the options are correct and similarly, they can eliminate it by knowing if only one of the options is incorrect.

18. Do not use "none of the above" as the last option, as it only measures students' ability to detect incorrect answers. Furthermore, if "none of the above" is the correct option, the SME must be certain that there are no exceptions to any of the options that the candidate may detect.

19. Do not use fill-in-the-blank format whereby a word is omitted in the middle of a sentence and the student must guess the correct word. All options should be placed at the end of the stem.

APPENDIX D

QUICK REVIEW

<u>Saudi Commission for Health Specialties (SCHS) Medical Licensing Examinations (SMLE)</u>

SMLE is a screen and as such not a certification of the original specialist qualification, but an honest attempt to identify those practitioners who do not meet a reasonable minimum competence to safely treat the Saudi public. Therefore, the exam shall focus on knowledge and competencies central to that particular specialty.

To achieve the goals of an SMLE, 50%-60% of items shall be problem solving using applied knowledge, 20%-30% interpretation of patient data and no more than 20% recall of factual knowledge with high relevance to the actual practice of medicine. It is expected that all candidates being granted a license to practice medicine in KSA shall also have demonstrated sufficient basic medical science underpinning medical practice to ensure that he is able to keep up with modern developments in his respective specialty.

Test Blueprints (TB) are provided for every SMLE and only items meeting the requirement laid down will be accepted. From time to time the TBs will be updated and therefore it is important to ensure that items are written according to the latest version.

SCHS policies and approach for writing high quality A-type multiple-choice items
General

1. Presentation of every MCQ shall be according to the in-house style:
    a. The vignette shall include in one paragraph the history and examination findings.
    b. If given, vital signs shall be presented in a tabulated format below the vignette.
    c. If given, lab data shall be presented in a tabulated format, with normal ranges and only SI units.
    d. If given, results of extra investigations, such as ultrasound, are appended in their own paragraph beneath the lab data, if given.
    e. A clear question line shall always be included in a paragraph of its own and a question mark at the end.
    f. There shall be four options, of which all are plausible, but only one is the best answer to the question asked.

2. The language used must be simple, and English jargon and colloquialisms avoided.
3. Abbreviations and acronyms shall be spelt out to avoid any confusion indirectly caused by the stress of the test.
4. Every item shall be checked for item writing flaws before submission.
5. All items shall be classified and references given.
6. Items shall be submitted in a template format for easy upload into the database used.

Content

1.   Questions shall address important practice issues and not trivia, medical history or research findings that have not become main stream and accepted in the major textbooks on the subject.
2.   Questions shall include basic clinical competencies involved in making a diagnosis, ordering and interpreting investigations, developing management plans and administering treatment, where appropriate classifying disease and making prognoses, offering counseling and education related to the specialty of the SMLE.
3.   No question shall seek to trick a candidate, for example through the use of word formulations.
4.   In general, questions shall avoid addressing rare or obscure cases treated within the specialty, however, where such cases illustrate important specialty concepts and principles this is allowable.
5.   Question content shall be summarized to reduce unnecessary and redundant information which will be time consuming and non-contributory to read.
6.   Any clinical data shall be authentic, current and accurate.
7.   Do not base questions on protocol driven management which has not been well established.
8.   Only content that is aligned with the TB will be used in the SMLEs.
9.   Questions based on the same vignette must be clearly identified to prevent two or more questions that may cue each other being used in the same exam.
10.   Where a vignette is used, ensure the logical clinical note sequence is followed and avoid exceeding a total wpi = 90[8]. The maximum words for a complex vignette with vital signs, lab data and extra investigation results shall not exceed 100, while the total wpi < 150.

Hand Cover Test (HCT) – A SCHS policy

Every multiple choice item shall satisfy the hand cover test, unless the question line clearly states that the choices are those of the following options. This implies that the best answer is not necessarily the best in the realm of medicine, but of those alternatives given it can be argued, motivated and justified as being the best of the choices offered.

Otherwise, the overwhelming majority of questions shall satisfy the HCT. This means that a candidate who covers the options can still answer the question correctly using the information in the vignette. Or where a vignette is not used, usually a recall type question, the single best / correct answer will be found among the four options offered.

Options and Distractors

1.   All SMLEs have four options.
2.   Option length is important and shall be approximately half to three quarters of a line in length, and all options shall be the same length.
3.   To achieve equal option length the definite or indefinite articles will usually be omitted.
4.   Options will often appear in summary forms, but the meaning must remain clear.
5.   All options shall be homogeneous, or on the same continuum, for example, all diagnoses
6.   Distractors shall be plausible, and not added as *fillers*. Only four options are used in SCHS exams to reduce this likelihood.
7.   Options shall not be overlapping, and ranges shall usually be used for quantitive data.

---

[8] wpi, words per item

Item Writing Flaws (IWFs)

Items with the following IWFs will be rejected or require editing:

1. Questions with NOT, EXCEPT, LEAST or ONLY in the question line..
2. The question asked is unambiguous.
3. The options are a series of true and false statements.
4. Redundancy, particularly information in a vignette.
5. The item does not satisfy the SCHS in-house style.
6. Convergence or mutually exclusive cues among the options.
7. The same or similar words are repeated in the stem and correct keyed option.
8. Logical clues which lead to the keyed option.
9. Option not similar in length.
10. Absolute or vague terms qualifying option statements.
11. The fourth option is: *none of the above* or *all of the above*.
12. Interpreted data presented for the sake of giving a context to the item.

Images

SCHS exams shall use ~10% - 20% items with high quality images. Certain specialties that rely more on images (e.g. Radiology, Pathology,…etc.) will have a higher percentage of images that will be decided by the exam committee.

1. Ensure that the image quality is appropriate, enough to view on a 17 inch computer screen without any pixilation.
2. Refer to the image in an appropriate way and letters and arrows used with a quick reference guide language: what does **A** show? Not, what is the tip of the arrow pointing at?

Units and normal values

All lab values must include the units and the range of normal values presented in a tabulated form. All laboratory data shall be with the SI units.

PLEASE READ THROUGH ALL YOUR QUESTIONS AND MAKE SURE THEY SATISTFY THE ABOVE AND DO NOT FORGET TO INCLUDE THE SOURCE REFERENCES AND TB CLASSIFICATION

APPENDIX E

GLOSSARY
www.pearsonassessments.com/pai/ai/research/resources/Glossary.htm. last visited on 11/17/2011

Correlation: The degree of relationship between two sets of scores. A correlation of 0.00 denotes a complete absence of relationship. A correlation of plus or minus 1.00 indicates a perfect (positive or negative) relationship. Correlation coefficients are used in estimating test reliability and validity.

Difficulty Index: The percent of students who answer an item correctly, designated as *p.* (At times defined as the percent who respond incorrectly, designated as *q.*)

Discrimination Index: The extent to which an item differentiates between high-scoring and low-scoring examinees. Discrimination indices generally can range from -1.00 to +1.00. Other things being equal, the higher the discrimination index, the better the item is considered to be. Items with *negative* discrimination indices are generally items in need of rewriting.

Distracters: An incorrect choice in a multiple-choice or matching item (also called a foil).

Error of Measurement: The amount by which the score actually received (an observed score) differs from a hypothetical true score. (See also Standard Error of Measurement .)

Item Analysis: The process of examining students' responses to test items to judge the quality of each item. The difficulty and discrimination indices are frequently used in this process. (See Difficulty Index and Discrimination Index .)

Mean ($\overline{X}$): The arithmetic average of a set of scores. It is found by adding all the scores in the distribution and dividing by the total number of scores.

Median (Md): The middle score in a distribution or set of ranked scores; the point (score) that divides a group into two equal parts; the 50th percentile. Half the scores are below the median, and half are above it.

Mode: The score or value that occurs most frequently in a distribution.

N: The symbol commonly used to represent the number of cases in a group.

Objectives: Stated, desirable outcomes of education.

p-Value: The proportion of people in an identified norm group who answer a test item correctly; usually referred to as the difficulty index. (See Difficulty Index .)

Reliability: The extent to which test scores are consistent; the degree to which the test scores are dependable or relatively free from random errors of measurement. Reliability is usually expressed in the form of a reliability coefficient or as the standard error of measurement derived from it. The reliability of a major classroom achievement test should be at least .60. The reliability of a standardized achievement or aptitude test should be at least .85. The higher the reliability coefficient the better, because this means there are smaller random errors in the scores. A test (or a set of test scores) with a reliability of 1.00 would have a standard error of zero and thus be perfectly reliable. (See Standard Error of Measurement .)

Reliability Coefficients: Estimated by correlation between scores on two equivalent forms of a test, by the correlation between scores on two administrations of the same test, or through procedures known as internal-consistency estimates. Each of the three estimates pertains to a different aspect of reliability. One of the easier and more commonly used (by teachers) estimates of reliability is known as the Kuder-Richardson Formula #21 estimate. The formula is

$$KR_{21} = r_{xx} = \frac{n}{n-1}\left(1 - \frac{\overline{X}n - \overline{X}}{ns_x^2}\right)$$

where n = number of items in the test

$\overline{X}$ = mean of the test

s_x^2 = variance of the test

Standard Deviation (S.D.) A measure of the variability, or dispersion, of a distribution of scores. The more the scores cluster around the mean, the smaller the standard deviation. In a normal distribution of scores, 68.3% of the scores are within the range of one S.D. below the mean to one S.D. above the mean. Computation of the S.D. is based upon the square of the deviation of each score from the mean. One way of writing the formula is as follows:

$$S.D. = \sqrt{\frac{\sum\left(X - \overline{X}\right)^2}{N}} \quad \text{where } X = \text{raw score}$$

$\overline{X}$ = mean

N = number of students

Standard Error of Measurement (SEM): The amount an observed score is expected to fluctuate around the true score. For example, the obtained score will not differ by more than plus or minus one standard error from the true score about 68% of the time. About 95% of the time, the obtained score will differ by less than plus or minus two standard errors from the true score.

$$SEM = S.D. \sqrt{1 - r_{xx}} \quad \text{where } S.D. = \text{standard deviation and}$$

r_{xx} = estimated reliability

The SEM is frequently used to obtain an idea of the consistency of a person's score or to set a band around a score. Suppose a person scores 110 on a test where the S.D. = 20 and [equation] = .91. Then:

$$SEM = 20\sqrt{1 - .91} = 20\sqrt{.09} = 20(.3) = 6$$

We would thus say we are 68% confident the person's true score was between (110–1 SEM) and (110+1 SEM) or between 104 and 116.

True Score: A score entirely free of error; a hypothetical value that can never be obtained by testing, since a test score always involves some measurement error. A person's "true" score may be thought of as the average of an infinite number of measurements from the same or exactly equivalent tests, assuming no practice effect or change in the examinee during the testing. The standard deviation of this infinite number of scores is known as the standard error of measurement. (See Standard Error of Measurement .)

Validity: The extent to which a test does the job for which it is intended. The term validity has different connotations for different types of tests and, therefore, different kinds of validity evidence are appropriate for each.