

UNIVERSITY OF HAWAI'I AT HILO

Writing Multiple Choice and True/False Exam Questions

A good practice guide

January 2012



Developed by Siân Millard, UH Hilo Strategic Planning Coordinator, and Ben Chavez, Assistant Professor, Department of Pharmacy Practice. All rights reserved.

Contents

1.	Overview and Guide Objectives	3
2.	Purpose of Exams	3
3.	Characteristics of Exams	3
4.	Types of Exam Questions.....	4
5.	Blooms Taxonomy of Cognitive Levels: An Overview.....	4
6.	Multiple Choice Questions.....	5
	Structure of Multiple Choice Questions.....	5
	Strengths and Weaknesses of Multiple Choice Questions	6
	Developing Multiple Choice Questions	6
	Other Tips on Developing Response Options	8
	K-Type Questions	9
7.	True/False Questions	9
	Types of True/False Questions.....	9
	Strengths and Weaknesses of True/False Questions.....	10
	Developing True/False Questions.....	11
8.	Blooms Taxonomy of Cognitive Levels: Question Examples.....	11
	Tips for developing 'higher-order' MCQ's	13
9.	Pulling Everything Together: The Exam as a Whole.....	14
10.	Evaluating Exam Questions.....	14
11.	References.....	16
12.	Further Reading	16
	Annex I: Further Examples of 'Higher Order' MCQ's.....	17

Writing Multiple Choice and True/False Exam Questions:

A Good Practice Guide

1. OVERVIEW AND GUIDE OBJECTIVES

In this guide you will learn how to apply the art of question design to the development of effective multiple choice and true/false questions for exams that you set.

Specifically, this guide is designed to help you to:

- Understand the key purpose and characteristics of effective exams
- Understand the main types of open and closed exam questions and what they can be used for
- Understand the levels of *Bloom's Taxonomy of Cognitive Levels* and its application to exam question design
- Understand the structure and strengths and weaknesses of multiple choice and true/false questions
- Design effective multiple choice and true/false questions
- Present your overall exam in a clear and consistent format
- Evaluate the effectiveness of your exam questions and identify student learning needs from exam responses

2. PURPOSE OF EXAMS

The main purpose of exams is to test students' understanding of what has been taught in line with learning objectives. In that way, teachers can assess which students are performing well, which may require additional support, and which areas of study students are particularly struggling to grasp.

To design an exam you need to be very clear about what you want to test, and then write questions to meet that objective.

3. CHARACTERISTICS OF EXAMS

There are four key characteristics of effective exams. They need to be:

i. Valid

- The exam tests what you intend it to test
- It is consistent with the course/learning objectives

ii. Reliable

- Allows consistent measurement of student performance

- Discriminates between different levels of performance (i.e. you can determine from the exam results which students are doing well and which not so well in your class)

iii. *Recognizable*

- It is representative of what you taught in the class (i.e. it does not include anything that you have not mentioned in class or referred students to in homework)

iv. *Realistic*

- It is designed so that it can be completed in the time available

4. TYPES OF EXAM QUESTIONS

There are four main types of questions commonly used in exams:

- i. *Short answer*
 - ii. *Essay*
 - iii. *Multiple choice questions (MCQ's)*
 - iv. *True/false questions (strictly speaking these are a sub-type of MCQ's)*
- } Open questions
- } Closed questions

Short answer and essay questions invite an open response from students. They are commonly used to assess critical thinking and logical reasoning by inviting students to analyze, synthesize and evaluate. They challenge students to create rather than select a response. They however tend to be restricted to assessing a limited sample of the range of content from a course, and can be difficult and time consuming to grade consistently.

The focus of this guide is on the design of multiple choice and true/false questions. These questions invite students to select from a list of potential options, within which is the correct answer. There is a misconception that these types of questions are ‘easy’ and only test knowledge or comprehension rather than other higher forms of cognitive skills.

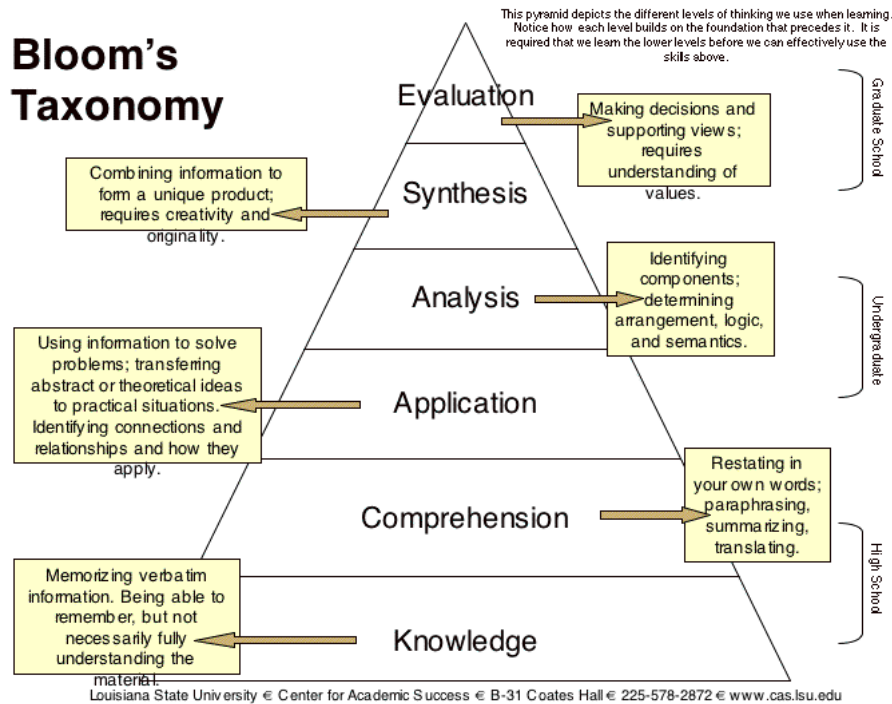
5. BLOOMS TAXONOMY OF COGNITIVE LEVELS: AN OVERVIEW

Benjamin Bloom created a Taxonomy of Cognitive Levels in 1956. The Taxonomy is a multi-tiered model of classifying thinking according to six levels. It provides a useful structure in which to categorize exam questions:

- Knowledge
 - Comprehension
 - Application
 - Analysis
 - Synthesis
 - Evaluation
- } Lower level
- } Upper level

The taxonomy is hierarchical – the first three (knowledge, comprehension and application) indicate ‘lower’ cognitive levels, and the last three (analysis, synthesis and evaluation) indicate ‘higher’

cognitive levels. The taxonomy can also be identified by what level of education each step typically covers (see illustration below):

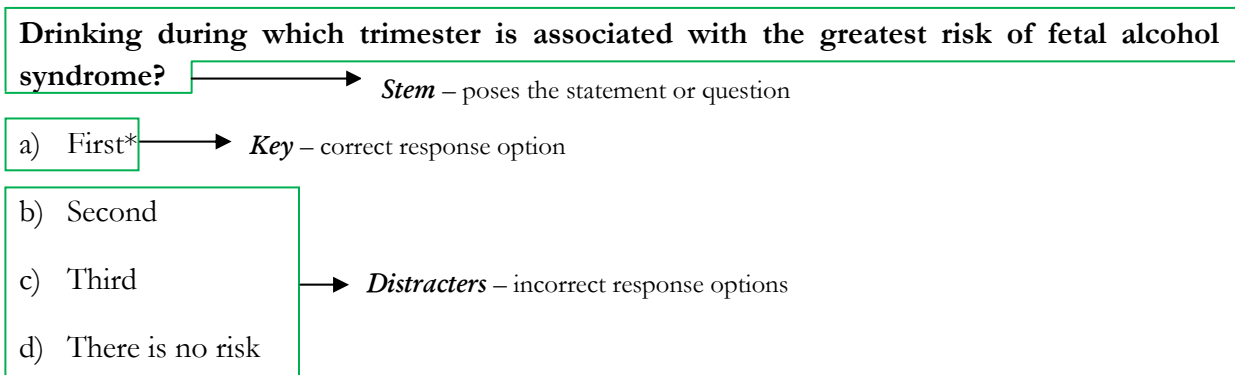


Multiple choice questions can be designed to test almost all of these different cognitive levels (apart from synthesis, see section 8). The type of question you ask will determine what cognitive level you test. Examples of different types of multiple choice and true/false questions by cognitive level are given in section 8.

6. MULTIPLE CHOICE QUESTIONS

Structure of Multiple Choice Questions

A multiple choice question (MCQ) has two core components – the stem and the response options. There are two types of response options – the key (correct option) and the distracters (incorrect options):



Stems can be written as a question or an incomplete statement which the response options then complete. In MCQ's there is typically only one correct response option out of a possible three to five. Research has shown that three-option items are as effective as four-choice options (see McKeachie (1986)), though note that students will have a greater probability of guessing the correct answer with three rather than four-options. Using more than five options makes it more difficult to come up with plausible distracters and it increases the reading load on students.

Strengths and Weaknesses of Multiple Choice Questions

MCQ Strengths	MCQ Weaknesses
Versatility in measuring different levels of cognitive skills	Can be difficult and time-consuming to construct properly
Permit a wide sampling of content and learning objectives	Ease of writing low-level knowledge items leads educators to neglect writing items to test higher-level thinking
Provide highly reliable test scores (objective)	Scores can be influenced by reading ability
Can be accurately and quickly scored (either by hand or by machine reader)	Can be difficult to find plausible distracters
Incorrect choices point to student learning needs	Does not measure writing ability
Reduced guessing factor compared with true/false items (because tends to be four response options rather than just two)	May encourage guessing (but less than true/false)

Developing Multiple Choice Questions

MCQ's should be developed and based on an educational or learning objective of the course. The tables below provide some tips on writing 'good' stems and response options, with examples where appropriate.

Examples (<i>correct responses are starred* where relevant</i>)		
Tips on developing STEMS	Not so good stem	Better stem
<p><i>Write items precisely, clearly, and as simply as possible</i></p> <ul style="list-style-type: none"> • don't be too wordy • don't include irrelevant information 	<p>Suppose you are a mathematics professor who wants to determine whether or not your teaching of a unit on probability has had a significant effect on your students. You decide to analyze their scores from a test they took before the instruction and their scores from another exam taken after the instruction. Which of the following t-tests is appropriate to use in this situation?</p> <p>a) Dependent samples*</p> <p>b) Heterogenous samples</p>	<p>When analyzing your students' pretest and posttest scores to determine if your teaching has had a significant effect, an appropriate statistic to use is the t-test for:</p> <p>a) Dependent samples*</p> <p>b) Heterogenous samples</p> <p>c) Homogenous samples</p> <p>d) Independent samples</p>

Examples (correct responses are starred* where relevant)		
Tips on developing STEMS	Not so good stem	Better stem
	c) Homogenous samples d) Independent samples	
<i>Be aware of cultural or language biases e.g. avoid colloquial phrases or terms that could be difficult for someone to understand where English is their second language</i>	Typically how old are keiki when they start elementary school? a) 3 years old b) 4 years old c) 5 years old d) 6 years old 'Keiki' is a Hawaiian term for children – literal translation meaning 'little one'. Not all students, especially if from out-of-state or international, may know this term.	Typically how old are children when they start elementary school? a) 3 years old b) 4 years old c) 5 years old d) 6 years old
<i>Avoid the use of negatives – or highlight them if you do – because they can be easily overlooked when reading</i>	Which of the following is not a required element for photosynthesis to occur? a) Carbon dioxide b) Oxygen* c) Sunlight d) Water	Which of the following is NOT a required element for photosynthesis to occur? a) Carbon dioxide b) Oxygen* c) Sunlight d) Water
<i>Ensure consistent grammar from the stem to all response options</i>	Albert Einstein was a: a) Anthropologist b) Astronomer c) Chemist d) Mathematician	Who was Albert Einstein? a) An anthropologist b) An astronomer c) A chemist d) A mathematician

Examples (correct responses are starred* where relevant)		
Tips on developing RESPONSE OPTIONS	Not so good response options	Better response options
<i>Keep a similar length for all response options</i> • Common mistake to write the correct option with more elaboration or detail	The term operant conditioning refers to the learning situation in which: a) A familiar response is associated with a new stimulus b) Individual associations are linked together in sequence c) A response of the learner is instrumental in leading to a subsequent reinforcing event* d) Verbal responses are made to verbal stimuli	The term operant conditioning refers to the learning situation in which: a) A familiar response is associated with a new stimulus b) Individual associations are linked together in sequence c) The learner's response leads to reinforcement* d) Verbal responses are made to verbal stimuli

Examples (correct responses are starred* where relevant)		
Tips on developing RESPONSE OPTIONS	Not so good response options	Better response options
<i>Ensure distracters are plausible</i>	Who was President of the United States during the War of 1812? a) Abraham Lincoln b) John F Kennedy c) James Madison* d) Richard Nixon	Who was President of the United States during the War of 1812? a) Thomas Jefferson b) James Monroe c) James Madison* d) John Quincy Adams
<i>Ensure response options are mutually exclusive</i>	What is the average effective radiation dose from a chest CT scan? a) 1-8 mS b) 8-16 mSv c) 16-24 mSv d) 24-32 mSv	What is the average effective radiation dose from a chest CT scan? a) 1-7 mSv* b) 8-15 mSv c) 16-24 mSv d) 25-32 mSv
<i>Reduce reading load by moving words into the stem where possible.</i>	Sociobiology can be defined as: a) the scientific study of humans and their environment b) the scientific study of animal societies and communication. c) the scientific study of etc.....	Sociobiology can be defined as the scientific study of: a) humans and their environment b) animal societies and communication c) etc.....
<i>Make sure there is only one correct or best response, or make clear that you are looking for the best answer (in which case all of the response options could be correct but one of them is the best).</i>	The function of the hypothesis in a research study is to provide a: a) tentative explanation of phenomena b) proven explanation of phenomena c) framework for interpretation of the findings d) direction for the research	According to the [lecture/author etc], the most important function of the hypothesis in a research study is to provide a: a) tentative explanation of phenomena b) proven explanation of phenomena c) framework for interpretation of the findings d) direction for the research

Other Tips on Developing Response Options

- Use common errors or misunderstandings to help develop distracters
- Place the response options in a logical sequence (E.g. alphabetical, numerical etc.)
- Use letters to label the response options (e.g. a), b), c) etc) instead of numbers. Numerical answers in numbered response options may be confusing to students.
- Avoid lifting text straight from the textbook
 - If it sounds like it right out of a textbook students are more likely to pick it out as the answer (they'll recognize different styles of writing).
- Use 'all of the above' carefully

- If ‘all of the above’ is the right answer, test-taker will likely select it if he/she recognizes two of the other three options as correct even if they don’t have a clue about the third.
- If the test-taker recognizes at least one incorrect response option then they can also eliminate ‘all of the above’, reducing their choice of options to two.
- Use ‘**none** of the above’ carefully
 - Use only when the correct answer is absolutely correct (e.g. in math).
 - If ‘none of the above’ is the correct answer, it won’t necessarily tell you if the student knows what the correct answer is – they may only recognize the incorrect answers.
 - Do not use in a negatively worded stem – double-negatives are confusing.
- Avoid using **don’t know** as a distracter as this is a waste of a response option and increases the likelihood of students guessing the correct answer.

K-Type Questions

You may have heard of K-Type questions (also known as combined-response or multiple multiple-choice questions) as a form of MCQ. In K-type questions one or more of the response options are correct answers and the remaining are distracters. The student is directed to identify the correct answer by selecting one of a set of letters, each of which represent a combination of alternatives.

Example:

The fluid imbalance known as edema is commonly associated with:

1. Allergic reactions
2. Congestive heart failure
3. Extensive burns
4. Protein deficiency

The correct answer is:

- a) 1, 2, and 3
- b) 1 and 3
- c) 2 and 4
- d) 4 only
- e) 1, 2, 3, and 4*

K-type questions have the disadvantage of providing clues that help students with only partial knowledge detect the correct combination of alternatives. In the example above, a student can identify combination *e* as the correct response simply by knowing that alternatives 1 and 4 are both correct. Because of this disadvantage, *k-type are not recommended.*

7. TRUE/FALSE QUESTIONS

Types of True/False Questions

With true/false questions, students decide whether a statement is true or false out of only two possible response options. These types of questions help to assess simple logic and can be used in

stating cause and effect relationships by using ‘because’ in the stem. As with multiple choice questions, there are two parts to a true/false question – the stem and the response options. One response option is the correct answer and the other is a distracter.

There are three types of true/false questions – (i) classic statement with true/false options, (ii) changing a false statement to make it true, and (iii) cause and effect.

i. Classic

Anabolic steroids provide an unfair advantage in sports where strength plays a large part

a) True*

a) False

ii. Changing a false statement to make it true

With this type of question, students indicate whether a statement is true or false; if false they must change an underlined word to make the statement true.

Subatomic particles of negatively charged electricity are called protons

a) True

b) False* *electrons*

iii. Cause and effect

For cause and effect true/false questions the stem has two parts (separated by the word ‘because’) and students must decide if the second part explains why the first part is true.

Leaves are essential *because* they shade the tree trunk

a) Yes

b) No*

Note that in this example true/false have been substituted for yes/no.

Strengths and Weaknesses of True/False Questions

True/False Strengths	True/False Weaknesses
Can be answered quickly by students (students can answer 50 true/false questions in the time it takes to answer 30 MCQs)	Susceptible to guessing (more than MCQ’s) as 50% chance of guessing the correct answer without any knowledge of content
Provide the widest sampling of content per unit of time	Can be difficult to write questions that are unequivocally true or false
Easy to score	Tend to be restricted to measuring knowledge and comprehension

Developing True/False Questions

True/false questions tend to be used to measure lower level cognitive skills – knowledge, comprehension and application. Many of the principles for developing MCQ's apply to the development of true/false questions. In summary, key tips for true/false questions are:

- Stems should be relatively short and simple
 - There is a tendency to add details to true stems which could be recognized by test-wise students
- Sweeping broad statements or absolutes tend to be false – students need only think of one untrue instance (e.g. **Students who make A's always have above average IQ scores**).
- Avoid ambiguous or vague statements (e.g. **A nickel is larger than a dime** – true if talking about diameter, but false if talking about monetary value)
- Arrange questions so that there is no discernible pattern of answers
- Response options can be True/False or Yes/No
- Avoid negatively worded stems or highlight negative words

8. BLOOMS TAXONOMY OF COGNITIVE LEVELS: QUESTION EXAMPLES

Bloom's Cognitive Level	Student Activity	Example Questions (<i>correct response options are starred*</i>)
Knowledge	Remembering facts, terms, concepts, definitions, principles	<p>Which one of the following persons is the author of "Das Kapital"?</p> <p>a) Mannheim b) Marx* c) Weber d) Engels</p> <p><i>[Note that the responses are internally consistent - they are all the names of Germans whose written work have been major contributions on social issues.]</i></p> <p>For a true/false question this could look like: Mannheim was the author of "Das Kapital"</p> <p>a) True b) False*</p> <p>or</p> <p>Which of the following is a likely side effect of Drug X:</p> <p>a) Diarrhea b) Constipation c) Tachycardia d) Elevated LFTs</p> <p>(Note that putting both diarrhea and constipation could potentially bias the</p>

		other answers, if a drug causes diarrhea it probably does not cause constipation)
Comprehension	Understanding the meaning of material	<p>If living cells similar to those found on earth were found on another planet where there was no molecular oxygen, which cell part would most likely be absent?</p> <p>a) Cell membrane b) Nucleus c) Mitochondria* d) Ribosome</p> <p>For a true/false question this could look like: If living cells similar to those found on earth were found on another planet where there was no molecular oxygen, mitochondria would most likely be absent</p> <p>a) True* b) False</p> <p>Drug X causes constipation due to its:</p> <p>a) blocking of muscarinic receptors b) agonism of alpha-2 receptors c) blocking of calcium channels d) blocking of 5HT1 receptors</p> <p>(note: having one answer drastically different from others may cause bias)</p>
Application	Using a concept or principle to solve a problem	<p>A school developed an aptitude test to use for admission to its Honors Program. The test was administered to a group of seven applicants who obtained the following scores: 70, 72, 72, 80, 89, 94 and 98. The mean score on the aptitude test is:</p> <p>a) 72 b) 82 * c) 80 d) 90</p> <p>For a true/false question this could look like: <i>A school developed an aptitude test to use for admission to its Honors Program. The test was administered to a group of seven applicants who obtained the following scores: 70, 72, 72, 80, 89, 94 and 98.</i></p> <p>The mean score on the aptitude test was 81</p> <p>a) True b) False*</p> <p>A 70 year old female patient, who is a smoker, and has controlled hypertension on metoprolol has a Framingham 10-year risk of:</p> <p>a) 5%</p>

		<p>b) 10%</p> <p>c) 20%</p> <p>d) 30%</p>
Analysis	Breaking material down into its component parts to see inter relationships/hierarchy of ideas	<p><i>It is said that when organic chemist Auguste Kekulé was struggling with how the six carbon atoms of benzene linked together, he dreamt of molecules twisting and turning around like snakes. In his dream, one of the snakes swallowed its own tail and rolled around like a hoop. When Kekulé woke up, he realized that the six carbon atoms of benzene were attached to each other to form a ring. Further work showed that this was correct.</i></p> <p>Which phase of the creative process is illustrated by this example?</p> <p>a) Preparation</p> <p>b) Incubation</p> <p>c) Orientation</p> <p>d) Illumination*</p> <p>A male patient presenting to his doctor's office with polyuria, polydipsia, and polyphagia would best be treated with:</p> <p>a) metformin</p> <p>b) aspirin</p> <p>c) sulfonylurea</p> <p>d) atorvastatin</p> <p>(Student needs to realize what disease state the patient is suffering from, then make a treatment recommendation based on that)</p>
Synthesis	Producing something new or original from component parts	<p>Note: There is debate over whether MCQ's can be used to test the synthesis level. That is because by its nature the synthesis level requires students to produce new or original material to answer the question. As MCQ's provide answer options there is no opportunity for something 'new' or 'original' to be developed as the answer will be within the provided list. It is therefore recommended that testing of the synthesis level is best achieved through short answer or essay type questions as they require the student to 'produce something new'.</p>
Evaluation	Making a judgment based on a pre-established set of criteria	<p>Disregarding the relative feasibility of the following procedures, which of these lines of research is likely to provide us with the most valid and direct evidence as to evolutionary relations among different species?</p> <p>a) Analysis of the chemistry of stored food in female gametes</p> <p>b) Analysis of the form of the Krebs cycle</p> <p>c) Observation of the form and arrangement of the endoplasmic reticulum</p> <p>d) Comparison of details of the molecular structure of DNA*</p>

Tips for developing 'higher-order' MCQ's

- Present students with a diagram of equipment and ask for application, analysis or evaluation
- Present actual quotations taken from newspapers or other published sources and ask for the interpretation or evaluation of these quotations

- Use pictorial materials that require students to apply principles and concepts
- Use charts, tables or figures that required interpretation
- Use case vignettes that require students to apply their knowledge

9. PULLING EVERYTHING TOGETHER: THE EXAM AS A WHOLE

- Consistent formatting of questions and response options will help students navigate through the exam
 - Clear instructions at start of exam – e.g. how long student has to complete questions, no calculator, no talking etc
 - Question numbers in bold
 - Stems in bold
 - Response options in normal font
 - Response options listed vertically rather than horizontally
 - Consistent identification of response options throughout (e.g. a, b, c, d or i, ii, iii, iv)
 - Consistent spacing between and within questions with more space between questions than within so that it is clear where one question ends and another begins
- Piloting is essential - enlist the help of colleagues to check questions for difficulty, ambiguity and accuracy.
- Group items by format or topic
 - E.g. all true/false together, all multiple-choice together – why? Reduces reading load on students
- Guard against order effects
 - Make sure the correct answer occurs as a) about the same amount as b) etc. Exam writers typically put the correct answer at either b) or c)
 - Ensure the stem of one question doesn't give away the answer to a different question

10. EVALUATING EXAM QUESTIONS

Evaluating your exam questions is an important exercise to:

- determine how well the exam measured what it intended to measure,
- determine whether there were any particularly problematic questions in terms of the way they were designed,
- identify student learning needs by identifying questions that were commonly answered incorrectly

To help inform your evaluation look at the distribution of correct and incorrect answers across questions and get student feedback in class discussion of the exams effectiveness – which questions did they find most difficult, which most easy to answer etc.

11. REFERENCES

Bloom, B.S. (1956) *Taxonomy of Educational Objectives, Vol 1*. New York: MacKay

McKeachie, W.J. (1986) *Teaching Tips 8th Ed* Lexington, Mass: Heath

12. FURTHER READING

Aiken, Lewis R., (1982). Writing multiple-choice items to measure higher-order educational objectives. *Educational and Psychological Measurement*, 1982, Vol. 42, pp. 803-806.

Burton, S.J., Sudweeks, R.R., Merrill, P.F., & Wood, B. (1991) *How to Prepare Better Multiple-Choice Test Items: Guidelines for University Faculty*. Accessed at <http://testing.byu.edu/info/handbooks/betteritems.pdf>, December 2011

Collins, J. (2006) Writing Multiple Choice Questions for Continuing Medical Education Activities and Self-Assessment Modules. *Radiographics: The journal of continuing medical education in radiology*, 26 (2), 543-551. Accessed at <http://radiographics.rsna.org/content/26/2/543.full.pdf+html>, December 2011

Marshall, J.C., & Hales, L.W. (1971) *Classroom Test Construction* Reading, MA: Addison-Wesley

Morrison, Susan, and Kathleen Walsh Free, (2001) Writing multiple-choice test items that promote and measure critical thinking. *Journal of Nursing Education*, January 2001, Vol. 40, No. 1, pp. 17-24.

Piontek, M.E. *Best Practices for Designing and Grading Exams* Center for Research on Learning and Teaching: University of Michigan, No.24. Accessed at http://www.crlt.umich.edu/publinks/CRLT_no24.pdf, December 2011

Zimmaro, D.M. (2004) *Writing Good Multiple-Choice Exams*. Accessed at <http://ctl.utexas.edu/assets/Evaluation--Assessment/Writing-Good-Multiple-Choice-Exams-04-28-10.pdf>, December 2011

ANNEX I: FURTHER EXAMPLES OF 'HIGHER ORDER' MCQ'S

Analysis

Mitochondria are called the powerhouses of the cell, because they make energy available for cellular metabolism. Which of the following observations is most cogent in supporting this concept of mitochondrial function?

- a) ATP occurs in the mitochondria.
- b) Mitochondria have a double membrane.
- c) The enzymes of the Krebs cycle, and molecules required for terminal respiration, are found in Mitochondria.
- d) Mitochondria are found in almost all kinds of plant and animal cells.
- e) Mitochondria abound in muscle tissue.

Evaluation

A student was asked the following question: "Briefly list and explain the various stages of the creative process."

As an answer, this student wrote the following:

"The creative process is believed to take place in five stages, in the following order: orientation, when the problem must be identified and defined, preparation, when all the possible information about the problem is collected, incubation, when no solution seems in sight and the person is often busy with other tasks, illumination, when the person experiences a general idea of how to arrive at a solution to the problem, and finally verification, when the person determines whether the solution is the right one for the problem."

How would you judge this student's answer?

- a) EXCELLENT (all stages correct in the right order with clear and correct explanations)
- b) GOOD (all stages correct in the right order, but the explanations are not as clear as they should be)
- c) MEDIOCRE (one or two stages are missing OR the stages are in the wrong order, OR the explanations are not clear OR the explanations are irrelevant)
- d) UNACCEPTABLE (more than two stages are missing AND the order is incorrect AND the explanations are not clear AND/OR they are irrelevant)