
Constructing Quality Multiple Choice Questions for Student Assessment

Contents

Introduction	1
MCQ – What is it?	1
Strengths and limitations of MCQ	2
Examples of forms of MCQs	3
General rules for constructing MCQs	5
Constructing MCQs for assessing complex achievement	13
Item analysis	21
Activity	23
References	25
Appendix 1 (Reading)	26

Introduction

Multiple choice questions (MCQs) provide faster ways of assessing student learning. MCQs are referred to as objective test items because they can be marked objectively. That is, different markers (and use of machine marking) can mark them independently and obtain the same results.

Multiple choice questions can be used effectively at almost all levels to measure a wide range of abilities. Well designed MCQs can measure higher end abilities such as analysis and evaluation.

MCQ – What is it?

The multiple choice question consists of a stem, which presents a problem situation, and several alternatives, which provide possible solutions to the problem. The stem may be a question or an incomplete statement. The alternatives include the correct answer and several plausible wrong answers, called distracters. The function of the distracters is to distract those students who are uncertain of the answer.

The following questions (also called items) illustrate the question form and the incomplete statement form.

1. What will you do first if an electric appliance in your office is on fire?
 - A. Dial 000
 - B. Report to the person concerned
 - C. Evacuate from the site at once
 - D. Switch off the power*

2. If an electric appliance in your office is on fire, the first thing you will do is to
 - A. dial 000.
 - B. report to the person concerned.
 - C. evacuate from the site at once.
 - D. switch off the power.*

In these examples both stems pose the same problem, but the incomplete statement is more concise. The above examples also illustrate the use of four alternatives. MCQs typically include either four or five choices. The larger number will reduce the student's chances of obtaining the correct answer by guessing. However, it is not easy for the teacher to present five plausible choices all the times. It is worth noting that by adding an obviously wrong answer merely to obtain five alternatives would not improve the item. There is no reason why the items in a given test should all have the same number of

alternatives. Some might contain four and some five, depending on the availability of plausible distracters.

Strengths and limitations of MCQ

Strengths

1. It offers more flexibility for assessing a diversity of content than can be garnered from other forms of assessment.
2. It allows for a precise interpretation for content validity.
3. Purely objective in scoring and testing.
4. Easy to be administered both by students and the teacher.
5. Student success does not depend on his/her writing skills.
6. Results can easily be compiled and analysed to determine:
 - patterns of student learning outcomes,
 - level of difficulty of questions,
 - usefulness of questions,
 - follow-up action required.
7. Can be used with all subject areas.

Limitations

1. It inhibits students from expressing creativity or demonstrating original and imaginative thinking.
2. Question design is restrictive, forcing students to fit their understanding into the designer's way of understanding a concept.
3. Success of question depends on suitability of distracters.
4. Longer reading time required and students with poor reading skills may be disadvantaged, especially under time limit.
5. Some students may guess at answers without understanding them.

6. Time consuming to design good questions. It is very easy to construct poor questions, bad questions may be much worse than other methods of assessing the same learning outcome.

Examples of forms of MCQs

1. Correct answer

- (i) What arithmetic process should be used when deciding how many cookies each child should get when distributing twelve cookies equally among three children?
- A. addition
 - B. subtraction
 - C. multiplication
 - D. division*
- (ii) What is the major advantage of a teacher-made test over a standardised test?
- A. It is usually an objective test thus eliminating subjective judgement.
 - B. It is usually better suited to the educational objectives of the class.
 - C. It is usually a nonverbal test, thus decreasing the reliance on language skills.
 - D. It gives the teacher an opportunity to discuss the items on the test with the students as they select their responses.*

2. Incomplete statement

- (i) The most significant use of language in the development of the human race is that it has become a means of
- A. preserving family unity.
 - B. developing the educational system.
 - C. creating national unity.
 - D. developing thinking in the individual.*
- (ii) An emotion can best be described briefly as
- A. a feeling of insecurity.
 - B. an attitude of aggression.
 - C. a stirred-up state of the organism.*
 - D. an elated state of the organism.

3. Combined response

- (i) A basketball falls freely from rest and hits the ground. It then rebounds to $\frac{1}{4}$ of its original height. Neglecting air resistance, which of the following statements about the basketball is/are correct?
- (1) Its kinetic energy just before collision is four times its kinetic energy just after collision.
 - (2) Its potential energy just before collision is four times its potential energy just after collision.
 - (3) The speed just before collision is two times the speed just after collision.
- A. (1) only.
 - B. (2) only.
 - C. (2) and (3) only.
 - D. (1) and (3) only.*
 - E. (1), (2) and (3).
- (ii) Which of the following capacitors are suitable for use in a high frequency circuit?
- (1) air capacitor,
 - (2) mica capacitor,
 - (3) ceramic capacitor,
 - (4) electrolytic capacitor.
- A. (1), (2) and (3).*
 - B. (1), (2) and (4).
 - C. (1), (3) and (4).
 - D. (2), (3) and (4).

4. 'Best answer'

- (i) Which one of the substances listed below is the poorest conductor of heat?
- A. Air*
 - B. glass
 - C. brick
 - D. water
- (ii) Which of the following has the greatest current when it is operated at 200V?
- A. a lamp with a resistance of 400Ω
 - B. a rice-cooker with rating of 400W at 200V
 - C. an electric iron with rating of 400W at 220V
 - D. a motor with a resistance of 250Ω

- E. a hair-dryer with rating of 600W at 200V

General rules for constructing MCQs¹

Ideally, a MCQ presents students with a task that is both important and clearly understood, and one that can be answered correctly only by those who have achieved the desired learning. The following rules for constructing MCQs are intended as general guides for preparing questions that hope to approximate this ideal.

Each question should be designed to assess an important learning outcome

The problem situation around which a question is to be built should be important and should be directly related to the learning outcomes (objectives) of the subject. Avoid for testing unimportant details, unrelated bits of information, and material that is irrelevant to the desired outcomes. The sample questions above illustrate some of the more fundamental aspects of knowledge that might be assessed. In testing these knowledge outcomes, focus on the more important terms, facts, and principles. Do not increase question difficulty by resorting to the more obscure and less significant questions of knowledge. Remember each question is expected to call forth student response that will help determine the extent to which the learning objectives of the subject have been achieved.

Present a single clearly formulated problem in the stem of the question

The task set forth in the stem should be so clear that a student can understand it without reading the alternatives.

Example

Poor:

A table of specifications:

- A. indicates how a test will be used to improve learning.
- B. provides a more balanced sampling of content.*
- C. arranges the instructional objectives in order of their importance.
- D. specifies the method of scoring to be used on a test.

Better:

What is the main advantage of using a table of specifications when preparing an achievement test?

- A. It reduces the amount of time required.
- B. It improves the sampling of content.*
- C. It makes the construction of test questions easier.
- D. It increases the objectivity of the test.

¹ This section is adapted from the “Rules for Constructing Multiple-Choice Items” section of Ch 3 of Gronlund, N.E. 1977. *Constructing Achievement Tests*. N.J. Prentice Hall.

State the stem in simple and clear language

The problem in the stem of a MCQ should be stated as precisely as possible and should be free of unnecessarily complex wording and sentence structure. Anyone who possesses the knowledge measured by a test question should be able to select the correct answer. Poorly stated stems frequently introduce sufficient ambiguity to prevent a knowledgeable student from responding correctly. Complex sentence structure may make the question a measure of reading comprehension than of the intended knowledge outcome.

Example

Poor: The paucity of plausible, but incorrect, statements that can be related to a central idea poses a problem when constructing which one of the following types of test questions?

- A. Short-answer.
- B. True-false.
- C. Multiple-choice.*
- D. Essay.

Better: The lack of plausible, but incorrect, alternatives will cause the greatest difficulty when constructing:

- A. short-answer questions.
- B. true-false questions.
- C. multiple-choice questions.*
- D. essay questions.

Another common fault in stating MCQs is to load the stem with irrelevant and, thus, nonfunctioning material.

Example

Poor: Testing can contribute to the instructional program of the school in many important ways. However, the main function of testing in teaching is:

Better: The main function of testing in teaching is:

Place as much of the wording as possible in the stem

Avoid repeating the same material in each of the alternatives. By moving all of the common content to the stem, it is usually possible to clarify the problem further and to reduce the time the student needs to read the alternatives.

Example

- Poor: In objective testing, the term objective:
- A. refers to the method of identifying the learning outcomes.
 - B. refers to the method of selecting the test content.
 - C. refers to the method of presenting the problem.
 - D. refers to the method of scoring the answers.*

- Better: In objective testing, the term objective refers to the method of:
- A. identifying the learning outcomes.
 - B. selecting the test content.
 - C. presenting the problem.
 - D. scoring the answers.*

Sometimes the problem is not simply to move the common words to the stem, but to reword the entire question.

Example

- Poor: Instructional objectives are most apt to be useful for test-construction purposes when they are stated in such a way that they show:
- A. the course content to be covered during the instructional period.
 - B. the kinds of behaviour students should demonstrate upon reaching the goal.*
 - C. the things the teacher will do to obtain maximum student learning.
 - D. the types of learning activities to be participated in during the course.

- Better: Instructional objectives are most useful for test-construction purposes when they are stated in terms of:
- A. course content.
 - B. student behaviour.*
 - C. teacher behaviour.
 - D. learning activities.

Whenever possible, state the stem in positive form

A positively phrased test question tends to measure more important learning outcomes than a negatively stated one. This is because knowing such things as the best method or the most relevant argument typically has greater educational significance than knowing

the poorest method or the least relevant argument. The use of negatively stated question items results all too frequently from the ease with which such questions can be constructed, rather than from the importance of the learning outcomes measured.

Example

- Question 1 Which one of the following is a category of the cognitive domain in Bloom's Taxonomy?
- A. Comprehension*
 - B. (distracter needed)
 - C. (distracter needed)
 - D. (distracter needed)
- Question 2 Which one of the following is **not** a category of the cognitive domain in Bloom's Taxonomy?
- A. Comprehension
 - B. Application
 - C. Analysis
 - D. (answer needed)

Note that Question 2 looks more easy to construct but it is apt to have a lower level of difficulty and is likely to measure relatively unimportant learning outcomes. Being able to pick the right answer that does not apply provides no guaranty that the student possesses the desired knowledge.

Whenever negative wording is used in the stem, emphasise it

Sometimes the use of negative wording is basic to the assessment of an important learning outcome. For an example, students have to know that certain chemicals should not be mixed in a chemistry lesson. When negative wording is used in the stem, it should be emphasised by using bold or upper cases and being placed near the end of the statement.

Example

- Poor: Which one of the following is not a desirable practice when preparing MCQs?
- A. Starting the stem in positive form.
 - B. Using a stem that could function as a short-answer question.
 - C. Underlining certain words in the stem for emphasis.
 - D. Shortening the stem by lengthening the alternatives.*

- Better: All of the following are desirable practices when preparing MCQs EXCEPT:
- A. stating the stem in positive form.
 - B. using a stem that could function as a short-answer question.
 - C. underlining certain words in the stem for emphasis.
 - D. shortening the stem by lengthening the alternatives.*

Make certain that the intended answer is correct or clearly best

When the correct-answer form of MCQ is used, there should be only one correct answer and it should be unquestionably correct. It may also be necessary to include 'of the following' in the stem to allow for equally satisfactory answers that have not been included in the question.

Example

Poor: What is the best method of selecting subject content for test questions?

Better: Which one of the following is the best method of selecting subject content for test questions?

To avoid equivocal answers when using the correct-answer form, proper phrasing of the stem can be helpful. An inadequately stated problem frequently makes the intended answer only partially correct or makes more than one alternatives suitable.

Example

Poor: What is the purpose of classroom testing?

Better: The main purpose of classroom testing is:

Check all alternatives are grammatically consistent with the stem and parallel in form

Grammatical inconsistency in tense, article or form could provide a clue to the correct answer, or at least make some of the distracters ineffective.

Example

Poor: The recall of factual information can be measured best with a:

- A. matching question.
- B. multiple-choice question.

- C. short-answer question.*
- D. essay question.

Better: The recall of factual information can be measured best with:

- A. matching questions.
- B. multiple-choice questions.
- C. short-answer questions.*
- D. essay questions.

Stating all of the alternatives in parallel form also prevents unnecessary clues to be given to students.

Example

Poor: Why should negative terms be avoided in the stem of a MCQ?

- A. They may be overlooked.*
- B. The stem tends to be longer.
- C. The construction of alternatives is more difficult.
- D. The scoring is more difficult.

Better: Why should negative terms be avoided in the stem of a MCQ?

- A. They may be overlooked.*
- B. They tend to increase the length of the stem.
- C. They make the construction of alternatives more difficult.
- D. They may increase the difficulty of the scoring.

Avoid verbal clues

Example 1 Similarity of wording in both the stem and the correct answer

Poor: Which one of the following would you consult first to locate research articles on achievement testing?

- A. Journal of Educational Psychology
- B. Journal of Educational Measurement
- C. Journal of Consulting Psychology
- D. Review of Educational Research*

Example 2 Stereotyped correct answer

Poor: Learning outcomes are most useful in preparing tests when they are:

- A. clearly stated in behavioural terms.*
- B. developed co-operatively by teachers and students.

- C. prepared after the instruction has ended.
- D. stated in general terms.

Example 3 Too much details in the correct answer may provide a clue

Poor: Lack of attention to learning outcomes during test preparation:

- A. will lower the technical quality of the questions.
- B. will make the construction of test questions more difficult.
- C. will result in the greater use of essay questions.
- D. may result in a test that is less relevant to the instructional program.*

Example 4 Use of two responses that are all inclusive makes it possible to eliminate the other alternatives.

Poor: Which one of the following types of test questions assesses learning outcomes at the recall level?

- A. Supply-type questions.*
- B. Selection-type questions.
- C. Matching questions.
- D. MCQs.

Example 5 Use of two responses that have the same meaning makes it possible to eliminate them as potential answers.

Poor: Which one of the following is the most important characteristic of achievement-test results?

- A. Consistency.
- B. Reliability.
- C. Relevance.*
- D. Objectivity.

Make the distracters plausible and attractive

The distracters should be so appealing to the students who lacks the knowledge required by the question that he/she selects one of the distracters in preference to the correct answer. Use of more homogeneous alternatives result in the greater plausibility.

Example

Poor: Obtaining a dependable ranking of students is of major concern when using:

- A. norm-referenced summative tests.*
- B. behaviour descriptions.

- C. check lists.
- D. questionnaires.

Better: Obtaining a dependable ranking of students is of major concern when using:

- A. norm-referenced summative tests.*
- B. teacher-made diagnostic tests.
- C. mastery achievement tests.
- D. criterion-referenced formative tests.

Vary the relative length of the correct answer to eliminate length as a clue

Example

Poor: One advantage of MCQs over essay questions is that they:

- A. measure more complex outcomes.
- B. depend more on recall.
- C. require less time to score.
- D. provide for a more extensive sampling of course content.*

Better: One advantage of MCQs over essay questions is that they:

- A. provide for the measurement of more complex learning outcomes.
- B. place greater emphasis on the recall of factual information.
- C. require less time for test preparation and scoring.
- D. provide for a more extensive sampling of course content.*

Avoid using the alternative ‘all of the above’ and use ‘none of the above’ with extreme caution

The use of ‘all of the above’ as an option makes it possible to answer the question on the basis of partial information. When ‘none of the above’ is used as the right answer in a correct-answer type of question, this option may be measuring nothing more than the ability to detect incorrect answers.

Example

Poor: Which one of the following is a category of the cognitive domain in Bloom’s Taxonomy?

- A. Critical thinking.
- B. Scientific thinking.
- C. Reasoning ability.
- D. None of the above.*

Vary the position of the correct answer in a random manner

The correct answer should appear in each alternative position about the same number of times, but this should not form a fixed pattern that may be apparent to the students.

Constructing MCQs for assessing complex achievement

MCQs are useful formats for assessing complex achievement. The rules of construction are the same as above described. Of course, greater skill is required in identifying and defining the problem situations. Also, the increased complexity of the tasks presented demands more careful phrasing of the questions.

To demonstrate the variety of complex learning outcomes that can be assessed, the major categories of intellectual skills and abilities listed in the cognitive domain of the Blooms' Taxonomy of Educational Objectives are described here first.

Bloom's Taxonomy of Cognitive Levels*Knowledge*

Recalling memorised information. May involve remembering a wide range of material from specific facts to complete theories, but all that is required is the bringing to mind of the appropriate information. Represents the lowest level of learning outcomes in the cognitive domain.

Learning objectives at this level: know common terms, know specific facts, know methods and procedures, know basic concepts, know principles.

Question verbs: Define, list, state, identify, label, name, who? when? where? what?

Comprehension

The ability to grasp the meaning of material. Translating material from one form to another (words to numbers), interpreting material (explaining or summarising), estimating future trends (predicting consequences or effects). Goes one step beyond the simple remembering of material, and represent the lowest level of understanding.

Learning objectives at this level: understand facts and principles, interpret verbal material, interpret charts and graphs, translate verbal material to mathematical formulae, estimate the future consequences implied in data, justify methods and procedures.

Question verbs: Explain, predict, interpret, infer, summarise, convert, translate, give example, account for, paraphrase x ?

Application

The ability to use learned material in new and concrete situations. Applying rules, methods, concepts, principles, laws, and theories. Learning outcomes in this area require a higher level of understanding than those under comprehension.

Learning objectives at this level: apply concepts and principles to new situations, apply laws and theories to practical situations, solve mathematical problems, construct graphs and charts, demonstrate the correct usage of a method or procedure.

Question verbs: How could x be used to y ? How would you show, make use of, modify, demonstrate, solve, or apply x to conditions y ?

Analysis

The ability to break down material into its component parts. Identifying parts, analysis of relationships between parts, recognition of the organisational principles involved.

Learning outcomes here represent a higher intellectual level than comprehension and application because they require an understanding of both the content and the structural form of the material.

Learning objectives at this level: recognise unstated assumptions, recognises logical fallacies in reasoning, distinguish between facts and inferences, evaluate the relevancy of data, analyse the organisational structure of a work (art, music, writing).

Question verbs: Differentiate, compare / contrast, distinguish x from y , how does x affect or relate to y ? why? how? What piece of x is missing / needed?

Synthesis

(By definition, synthesis cannot be assessed with multiple-choice questions. It appears here to complete Bloom's taxonomy.)

The ability to put parts together to form a new whole. This may involve the production of a unique communication (theme or speech), a plan of operations (research proposal), or a set of abstract relations (scheme for classifying information). Learning outcomes in this area stress creative behaviours, with major emphasis on the formulation of new patterns or structure.

Learning objectives at this level: write a well organised paper, give a well organised speech, write a creative short story (or poem or music), propose a plan for an experiment, integrate learning from different areas into a plan for solving a problem, formulate a new scheme for classifying objects (or events, or ideas).

Question verbs: Design, construct, develop, formulate, imagine, create, change, write a short story and label the following elements:

Evaluation

The ability to judge the value of material (statement, novel, poem, research report) for a given purpose. The judgments are to be based on definite criteria, which may be internal (organisation) or external (relevance to the purpose). The student may determine the criteria or be given them. Learning outcomes in this area are highest in the cognitive hierarchy because they contain elements of all the other categories, plus conscious value judgments based on clearly defined criteria.

Learning objectives at this level: judge the logical consistency of written material, judge the adequacy with which conclusions are supported by data, judge the value of a work (art, music, writing) by the use of internal criteria, judge the value of a work (art, music, writing) by use of external standards of excellence.

Question verbs: Justify, appraise, evaluate, judge *x* according to given criteria. Which option would be better/preferable to party *y*?

Comprehension questions²

Comprehension can be measured by requiring students to translate material from one form to another, to interpret the meaning of material, or to extrapolate – that is, extend the meaning beyond the data presented.

Example

Outcome: Ability to identify the meaning of a term (Translation)

² Examples used in this section are adapted from Gronlund, *op cit*.

1. Which one of the following is closest in meaning to the term *taxonomy*?
 - A. Classification.*
 - B. Construction.
 - C. Evaluation.
 - D. Translation.

2. Which one of the following true-false statements contains a *specific determiner*?
 - A. Australia is bigger than Europe.
 - B. Australia was discovered by the British.
 - C. Australia has some big primary industries.*
 - D. Australia's population is increasing.

Example

Outcome: Ability to grasp the meaning of an idea (Interpretation)

1. The statement 'Test reliability is a necessary but not a sufficient condition of test validity' means that
 - A. a reliable test will have a certain degree of validity.
 - B. a valid test will have a certain degree of reliability.*
 - C. a reliable test may be completely invalid and a valid test completely unreliable.

Example

Outcome: Ability to identify the explanation of a phenomenon (Interpretation)

1. If a true-false test is to replace a multiple-choice test, the number of questions will have to be greater. Why?
 - A. To maintain the same reliability.*
 - B. To maintain the same level of difficulty.
 - C. To maintain the same coverage of subject matter.
 - D. To allow for a larger number of true statements than false statements.

Example

Outcome: Ability to predict the most probable effect of an action (Extrapolation).

1. What is most likely to happen to the reliability of the scores on a multiple-choice test in which the number of alternatives for each question is changed from four to five?
 - A. It will increase.*
 - B. It will decrease.
 - C. It will stay the same.

Note that these examples would represent the comprehension level only if the situations were new to the students. If the solutions have been encountered during lectures, the questions can only be classified under the knowledge domain.

Application questions³

Application questions also measure understanding, but at a higher level in the cognitive domain than that of comprehension. Students must demonstrate that they not only grasp the meaning of information but can also apply it to concrete situations that are new to them.

Example

Outcome: Ability to apply facts and principles.

1. Which one of the following learning outcomes is properly stated in behavioural terms?
 - A. Student learns the purposes of achievement testing.
 - B. Student develops an appreciation of objective testing.
 - C. Student understands the advantages of using a table of specifications.
 - D. Student identifies errors in test questions.*

Example

Outcome: Ability to apply rules of procedure.

Read the following test question and then indicate the best change to make to improve the question.

1. Which one of the following types of learning outcomes is most difficult to evaluate objectively?
 1. A concept.
 2. An application.
 3. An appreciation.
 4. None of the above.

³ *Ibid.*

The best change to make in the above question would be to:

- A. change the stem to incomplete-statement form.
- B. use letters instead of numbers for each alternative.
- C. remove the indefinite article 'a' and 'an' from the alternatives.
- D. replace 'None of the above' with 'An interpretation.'*

Analysis questions⁴

Analysis includes identification of parts, relationship between parts, and the way the elements are organised.

Example

Outcome: Ability to recognise unstated assumptions.

1. The author of a recent book on measurement claims that students will obtain correct answers on 50% of the items in a true-false test by guessing. Which one of the following assumptions is the author making?
 - A. Students will make some informed guesses.
 - B. Students will guess blindly.*
 - C. Students' guessing will be aided by some specific determiners.
 - D. Students' guessing will be aided by both partial information and some specific determiners.

Outcome: Identifying relations between ideas.

1. Read the following two statements and select the answer that best explains their relationship.
 - One: Other things being equal, a longer test is more reliable than a shorter test.
 - Two: Other things being equal, a fifty-question multiple-choice test is more reliable than a fifty-question true-false test.

The relationship between these statements can best be expressed as follows:

 - A. The situation in statement two contradicts the principle in statement one.

⁴ *Ibid.*

- B. The situation in statement two can be explained by the principle in statement one.*
- C. The situation in statement two neither contradicts nor can be explained by the principle in statement one.

Example

Outcome: Ability to recognise forms, pattern, or structure that is implicit in some works.

1. Which one of the following statements best expresses the point of view of the author of this booklet towards MCQ testing?
 - A. Objectivity should be maintained even if some of the learning outcomes are measured less directly.
 - B. Learning outcomes should be measured directly even if less objective measures must be used.*
 - C. Neither measuring the learning outcomes directly nor maintaining objectivity is as important as following closely the rules for constructing test questions.

Critical thinking questions

Example

Outcome: Use of case study can provide material for several follow-up questions that demand critical thinking ability to answer.

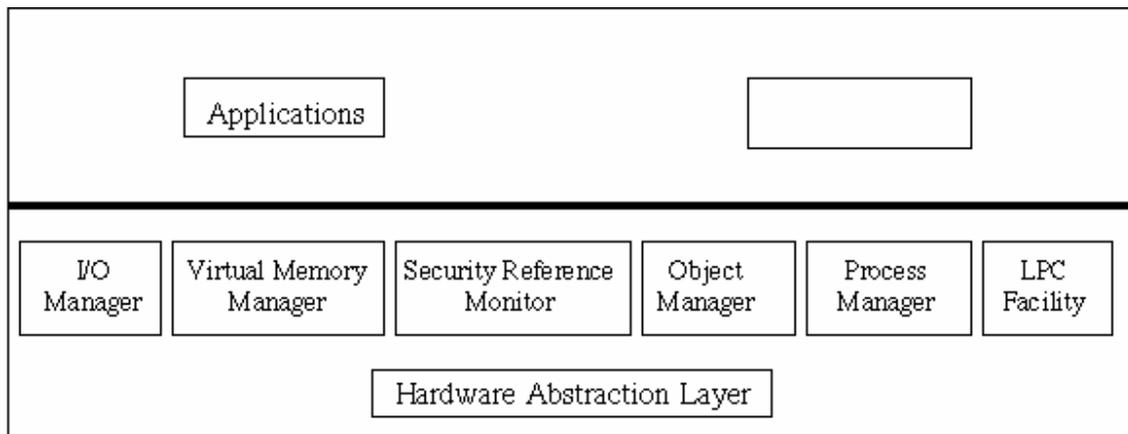
1. Alice, Barbara, and Charles own a small business: the Chock-Full-o-Goodness Cookie Company. Because Charles has many outside commitments and Barbara has a few, Alice tends to be most in touch with the daily operations of Chock-Full-o-Goodness. As a result when financial decisions come down to a vote at their monthly meeting, they have decided that Alice gets 8 votes, Barbara gets 7, and Charles gets 2-with 9 being required to make the decision. According to minimum-resource coalition theory, who is most likely to be courted for their vote?
 - A. Alice.
 - B. Barbara.
 - C. Charles.
 - D. No trend toward any specific person.
2. In the above scenario, according to minimum-power coalition theory, who is most likely to be courted for their vote?
 - A. Alice.
 - B. Barbara.

- C. Charles.
- E. No trend toward any specific person.

Example

Outcome: To use new form of language to present a familiar concept and to prevent students from using rote memorisation to answer the question.

1. Use the diagram below to answer the following questions.



- a) What belongs in the empty box in the upper right corner of the diagram?
 - A. Hardware devices.
 - B. Client Services for Netware.
 - C. Logon Process.
 - D. Gateway Services for Netware.

- b) If the Applications resided below the heavy black line, they would:
 - A. be open to hackers on the network.
 - B. compete with the OS for memory.
 - C. be preemptively multi-tasked.
 - D. launch in individual NTVDMs.

Item analysis

Item difficulty (facility)

A simple statistic which shows how easy or difficult an item is. It is the proportion of students getting an item correct.

$$\text{Facility} = \frac{\text{No. of correct response}}{\text{Total no. of students}}$$

The facility should preferably be around 0.6 (ie the expected mean for the class on the test is about 60%).

Item discrimination

Shows the degree to which a particular item discriminates between the able and less able students. It is an indication for the quality of the item. Assuming that students' total scores in a test and their scores on each item are positively and substantially correlated,

$$\text{Discrimination index} = \frac{(\text{No. correct in higher 27\% group}) - (\text{No. correct in lower 27\% group})}{\text{No. in higher group} - \text{No. in lower group}}$$

The index lies between +1 and -1. With an index of +1, there is a perfect relationship between students' test score and their scores on that item. If the index is greater than zero and less than 1, it indicates students with higher test scores tend to perform better on that item than those with less scores. If the index is equal to zero, it means all students answered the item correctly. If the index is a negative number, it means students with lower test scores are more likely to answer that item correctly than those with higher scores. The following are interpretations of some index values.

Index value	Interpretation
0.4 or over	Very good item
0.30 to 0.39	Reasonably good item, but can be improved
0.20 to 0.29	Marginal item that should be improved
Below 0.2	Poor item which <ul style="list-style-type: none"> • should be discarded unless it can be restructured. • Is either too difficult or too easy that it fails to discriminate high performers with low performers.

Distracter analysis

The discrimination index may help to indicate there is possibly a problem with an item, but the index itself does not reveal the source of the problem. A close examination of the content of the distracters would explain why high performers and low performers are attracted by a wrong answer. For example, look at the following case.

Distribution of alternatives selected for Question 1 (100 students)					
Student group	No. of students selecting each alternative				
	A	B	C	D	E*
Upper 27%	1	1	8	1	16
Middle 46%	1	2	9	6	30
Lower 27%	4	3	5	4	9
Total	6	6	22	11	55

E* - best/correct answer

- i) Data shows only 16 of the upper group answered the item correctly, but 30 in the middle group also got the same answer. Is this against our assumption about discrimination?
- ii) Most of the others in the upper group selected C. Is C too obvious than A, B and D?
- iii) Alternatives A and B are not very attractive to students. Are they too easy or too obvious that students can discern them from other alternatives? Should they be restructured or replaced?

The discrimination R biserial (also known as the point-biserial correlation coefficient)

This coefficient correlates the scores obtained on a particular item with the scores obtained on the test as a whole. High values indicate that the students who performed well on the whole test also performed well on that item and that students who performed badly on the whole test also performed badly on that question. Interpretation of this statistic remains a matter for your judgement, just as with the discrimination index.

Guessing corrections

Guesses may be intelligent, they may be based on partial information, mis-information or sheer ignorance. The formula to correct these forms make many assumptions. The standard formula assumes all wrong responses are of the same value. It over-corrects for guessing by weak students and under-corrects when the distracters are obvious. Guessing corrections do not usually change the rank order or distribution of scores. The formula is

$$CR = R - W / (K - 1)$$

Where CR = responses corrected for guessing

R = number of right responses

W = number of wrong responses

K = number of alternatives per item.

Activity

Select a topic from one of your subjects and construct four MCQs of four-choice alternatives. Vary the difficulty of the MCQs by including one question that requires only simple recall. One that assesses students' comprehension and the other two measures students application and critical thinking abilities respectively. Discuss the questions with your colleagues in this Workshop.

MCQ 1

MCQ 2

MCQ 3

MCQ 4

References

Gronlund, N. E. 1977. *Constructing Achievement Tests*. Englewood, Prentice-Hall Inc.

Isaacs, G. 1984. *Multiple Choice Testing: A Guide to the Writing of Multiple Choice Tests and to their analysis*. NSW, Higher Education Research and Development Society of Australasia.

Kehoe, J. 1995. 'Basic Item Analysis for Multiple Choice Tests' in *Practical Assessment, Research and Evaluation*. Vol 4 No 10.

Wiersma, W. and Jurs, S.G. 1990. *Educational Measurement and Testing*. Boston, Allyn and Bacon.

Appendix

Reading: Kehoe, J. 1995. 'Basic Item Analysis for Multiple Choice Tests' in *Practical Assessment, Research and Evaluation*. Vol 4 No 10.