

CAD Guidelines

Improving Learning and Teaching

A guide for developing multiple choice and other objective style questions

2013

Centre for Academic Development



CAD Guidelines

Contents

Section	Page
1. <u>Purpose, Structure And Objectives</u>	2
2. <u>Writing Multiple Choice Questions</u>	4
The format of a multiple choice question	4
Twelve guidelines for writing good multiple choice questions	6
Strengths and limitations of multiple choice questions	27
3. <u>Other Objective Question Formats</u>	28
True/false questions	28
Multiple selection questions	28
Assertion-reason questions	29
Matching pairs questions	29
Sentence completion questions	30
4. <u>Writing Objective Questions To Test More Than Recall</u>	31
5. <u>Answers to Exercises</u>	37

Authors: Cedric Hall and Stephen Marshall

Published 2013: Centre for Academic Development, Victoria University of Wellington, New Zealand.

CAD acknowledge with gratitude the work of Professor Cedric Hall in creating the original material for this guideline document. The current guideline has been revised and extended in part by CAD staff and all errors or omissions are our responsibility.

This guide draws upon material contained in assessment modules developed in 1987-89 for the Assessment of Student Performance (ASP) Project. This project was originally funded by the Authority for Advanced Vocational Awards, under the co-directorship of Brad Imrie and Cedric Hall, with the aim of supporting the in-service training of polytechnic tutors in techniques of assessment. The copyright (1994) for the ASP modules is held by Cedric Hall who has given permission for the use of ASP material in this guide.

For further information, contact Emeritus Professor Cedric Hall, Victoria University of Wellington, Faculty of Education, PO Box 17-310 Karori, Wellington, New Zealand. Email: cedric.hall@vuw.ac.nz.

1. Purpose, Structure And Objectives

Purpose

The purpose of this guide is to help teachers at all levels of education develop well written *objective-style questions*, such as multiple choice, true/false, and matching-pairs questions. These questions are described as *objective* in style because they have only **one correct or best** answer as opposed to questions which ask students to give opinions, argue a case, solve a new problem, create an art object, etc., where a range of answers are possible and defensible.

Assessment at Victoria is based on six key principles outlined in the [Assessment Handbook](#) (section 1.3):

Validity

Assessment should be fit for purpose. Assessment tasks should therefore be appropriate for the level, content and learning objectives of the course and the graduate attributes of the programme and university. A valid task will be one that measures what it purports to assess.

Reliability

Assessment should provide an accurate and consistent measure of student performance. This involves both consistency in marking and the authenticity of student work.

Fairness and inclusivity

Assessment tasks should provide every student with an equitable opportunity to demonstrate their learning. Tasks should not discriminate against students on the basis of gender, race, ethnicity, religion, disability or political affiliation.

Contribution to learning

Assessment should be recognised as a learning activity. Assessment tasks should contribute to the development of skills and knowledge that can be applied within the course as well as in other contexts.

Manageability

Assessment tasks should be reasonable and practicable in terms of time and resources for both students and staff.

Transparency

The intention and practice of assessment should be clearly described to students and to other staff teaching in a course so that its benefits, purposes and procedures are understood by all parties, in the spirit of a teaching and learning partnership.

[Return to index](#)

Objective-style questions, when used as part of a balanced and varied assessment scheme, can support these principles. In particular, their use as self-tests (commonly online) can support students through immediate feedback (see [Guideline 6, page 14](#)).

Objective-style questions are frequently criticized because they too often focus on asking students to recall basic information and ideas rather than show their ability to think deeply and creatively about knowledge and solutions to problems. Section 4 of this guide provides suggestions for taking objective questions beyond the assessment of recall and basic understanding. The principle of validity means that objective-style questions should not normally comprise the only form of assessment used in a course.

Structure of the guide

The guide sets out information and guidelines under the following headings:

1. Purpose, structure and objectives
2. Writing multiple choice questions
3. Other objective question formats
4. Writing objective questions to test more than recall

Section 2 on writing multiple choice questions includes a list of the strengths and weaknesses of multiple choice questions. Exercises are provided to give teachers practice in developing and critiquing multiple choice questions.

Learning objectives

After reading this guide, you should be able to:

1. Write good multiple choice questions.
2. Identify common faults in poorly written multiple choice questions.
3. Write multiple choice and other objective-style questions that test higher thinking skills such as:
 - analysis of data or information
 - integration of concepts or ideas from different parts of a course
 - the transfer of ideas and concepts from previous learning to new learning in a different area
 - the solution of new or unfamiliar problems.

2. Writing multiple choice questions

THE FORMAT of a multiple choice question

Multiple choice questions are the most commonly used format for presenting objective-style questions. A multiple choice question consists of two parts: A stem and several options or alternatives. The stem usually asks a question or makes a statement. The student has to identify either the **correct** or the **best** option to go with the stem. The incorrect options are called *distractors* and the correct option is often referred to as the *key*.

Example: Choosing the **correct** option

Incorrect responses in a multiple choice question are called: **(stem)**

- A. Ratings. **(distractor)**
- B. Options. **(distractor)**
- C. Stems. **(distractor)**
- D. Distractors. * **(key)**

Example: Choosing the **best** option

Which of the following words is closest in meaning to the term taxonomy? **(stem)**

- A. Organization. **(distractor)**
- B. Analysis. **(distractor)**
- C. Classification. * **(key)**
- D. Revenue. **(distractor)**

If a test contains both of these question styles, it is a good idea to say so to students in the instructions at the beginning of the test.

Note that this guide provides four options in most of the examples shown. There is no hard and fast rule as to the best number of options, but most writers recommend three, four or five.

CAD Guidelines

Punctuation

Teachers are often inconsistent in the way they punctuate multiple choice questions. For simplicity, this guide uses the following rules even though they may not always be technically correct:

1. Begin the first word of each option with a capital letter.
2. End each option with a full stop.
3. In all other situations use standard punctuation rules.

» *Now do Exercise 1* «

Exercise 1

Select one topic you have taught during the last two months.

Write one multiple choice question of the correct option variety and one multiple choice question of the best option variety.

Keep these questions beside you as you progress through the guide. After the discussion of each guideline check your questions for the fault described. If necessary, amend your questions.

CAD Guidelines

TWELVE GUIDELINES for writing good multiple choice questions

Summary of the Twelve Guidelines

Guideline

- 1 **Make sure that the wording of your questions is suitable for the reading level of your students.**
Consider reading/language levels of students for whom the questions are intended, especially those whose first language is not English.
- 2 **Make your questions concise, clear and grammatically consistent.**
Keep the question's stem concise, giving only the necessary information. Options should avoid repeating words and be grammatically consistent.
- 3 **Write stems that clearly state the topic of the question.**
Stems should be clearly focussed. Leave no doubt about the question's intent.
- 4 **Only one option should be correct or best (according to experts in the field).**
Ensure other options can't be justified as correct.
- 5 **Avoid negative statements if possible. If you must use a negative, write the negative in capital letters or underline it. NEVER use double negatives.**
Negatives can easily be missed by students under pressure.
- 6 **Provide plausible distractors.**
Implausible distractors reduce the number of options. Well-chosen distractors can be helpful to pinpoint topics that confuse students.
- 7 **Don't give clues to the correct choice.**
Ensure a singular or plural match between the stem and all options. Avoid using words that denote frequency, such as 'always', 'never', 'sometimes' or 'often'. Use distractors of similar length to avoid over-qualifying the correct choice.
- 8 **Don't use overlapping alternatives.**
Ensure options are clearly different.
- 9 **Don't ask students to express an opinion.**
Avoid including the word 'you' in the stem, as this invites opinion. Essays are a better method for measuring opinion.
- 10 **NEVER use 'all of the above'. Be careful when using 'none of the above'.**
A student needs only partial information to choose either of those choices.
- 11 **Validate your questions after each use.**
Validation helps you to evaluate the effectiveness of your questions through analysis of students' responses.
- 12 **Always have your questions checked by a colleague.**
A colleague may find ambiguities, unintentional clues or obvious omissions in the test content that you have missed.

[Return to index](#)

CAD Guidelines

Explanation and examples of the Twelve Guidelines

Guideline 1 Make sure that the wording of your questions is suitable for the reading level of your students.

There is no point in using language that is beyond the reading or language level of your students. Your purpose is to test the ability of students to deal with ideas and perform skills within their chosen field, not their ability to cope with difficult language. This applies to any material written for students such as class notes, laboratory instructions, assignment information, and examination papers. Consider reading levels for students whose first language is not English.

Guideline 2 Make your questions concise, clear and grammatically consistent.

Example: Poor question

The use of multiple-choice questions expanded following the first world war because of their efficiency in helping recruitment and placement decisions within the military. Despite the fact that such questions are not usually suited to testing the ability of people to write clearly and argue a case, they are used widely in many educational contexts. Why?

- A. They are easy to construct.
- B. They are easy to score for large groups of students. *
- C. They eliminate issues related to the authenticity of students' work.
- D. Their convenience for distance and on-line learning.

In this example, the stem is too long-winded and gives too much useless information. The question need only ask why multiple questions are used widely in many educational contexts.

Note that the grammatical structure of the options is inconsistent. Options A, B and C are sentences while D is a phrase. Questions should be grammatically consistent.

Example: Better question

Multiple choice questions are used widely in many educational contexts because:

- A. They are easy to construct.
- B. They are easy to score for large groups of students. *
- C. They eliminate issues related to the authenticity of students' work.
- D. They are convenient for distance and on-line learning.

A further point is not to repeat words unnecessarily in each option. Take them out altogether, or put them in the stem.

Example: Poor question

The reliability of a test administered on one occasion only is usually estimated:

- A. By analysing data for its internal relevance.
- B. By analysing data for its internal consistency. *
- C. By analysing data for its external relevance.
- D. By analysing data for its external consistency.

Example: Better question

The reliability of a test administered on one occasion only is usually estimated by analysing data for its:

- A. Internal relevance.
- B. Internal consistency. *
- C. External relevance.
- D. External consistency.

Guideline 3 Write stems that clearly state the topic of the question.

Example: Poor example

Standards based assessment

- A. Is suited to all educational assessment purposes.
- B. Requires students to be assessed by standardised examinations.
- C. Has been criticised for controlling what is taught.
- D. Focuses on comparing students' answers with written criteria. *

In this exaggerated example, the weak stem fails to give the student a clear focus. The only clue is that the question deals with standards based assessment. The options span a wide range of possible themes that could be probed in relation to standards based assessment. The following example leaves the student in no doubt about the intent of the question.

Example: Better question

Which of the following features best describes standards based assessment?

- A. Students' performances are compared to pre-determined norms.
- B. The progress of students is tracked over time.
- C. Students' answers are compared to pre-set written criteria. *
- D. The achievement of students is profiled across different subjects.

CAD Guidelines

Guideline 4 Only one option should be correct or best (according to experts in the field).

Example: Poor question

What is the purpose of classroom testing?

- A. To diagnose learning difficulties.
- B. To help student learning by giving feedback.
- C. To grade student performance.
- D. To stream students by ability bands.

Depending on what your aim is, you could justify any of these options.

Example: Better question

The main aim of a diagnostic test is to:

- A. Stream students by ability bands.
- B. Provide feedback on overall progress to students.
- C. Identify specific strengths and weaknesses in a student's performance. *
- D. Motivate student learning.

Guideline 5 Avoid negative statements if possible. If you must use a negative, write the negative in capital letters or underline it. NEVER use double negatives.

Example: Poor question

Which of the following is not an expected feature of a good norm referenced test?

- A. Questions should range in difficulty.
- B. Results should not highlight individual differences. *
- C. Questions should produce high discrimination indices.
- D. Results should not correlate well with school achievement.

[Return to index](#)

CAD Guidelines

Under the pressure of testing and being familiar with the material, many students can miss a negative word such as “not” and “never” or a word with a negative prefix such as “uncharacteristic” and “unimportant”. The “not” in the stem of the above question comes into this category.

The question is made much worse by also having negatives in two of the options, B and D. Many students would be confused by the double negative caused by the “not” in the stem and the “not” in these options.

Example: Better question

Which of the following is a key characteristic of all norm referenced tests?

- A. It shows each student's mastery of the relevant educational standards.
- B. It identifies a person's standing relative to others. *
- C. It identifies areas for further revision.
- D. It identifies readiness for advanced study.

» ***Now do Exercise 2*** «
(starting on next page)

Exercise 2

Below you will see five multiple choice questions. Each question contains one or more of the faults discussed under guidelines 1 to 5.

Underneath each question, write down the fault(s) it contains. Use the following list to help you.

- a. Inappropriate language level (guideline 1).
- b. Unclear or ambiguous wording (guideline 2).
- c. Lack of conciseness (guideline 2).
- d. Grammatically incorrect (guideline 2).
- e. Poorly defined stem (guideline 3).
- f. More than one correct or best option (guideline 4).
- g. Double negative (guideline 5).
- h. Avoidable single negative (guideline 5).
- i. Single negative not emphasised (guideline 5).

1. Which is not a benefit of using a topic order diagram
- A. They allow tutors to plan the sequence of their course.
 - B. They allow tutors to work out alternative pathways through their course topics.
 - C. They allow tutors to develop alternative course objectives.
 - D. They allow tutors to work out points in the course sequence for placing assessments.

Fault(s) _____

2. Research has shown that
- A. Proportionately more people in the age range 16-25 are involved in fatal road accidents.
 - B. Against predictions, the rate of repossession of motor vehicles by loan companies has decreased in the past twelve months.
 - C. The use of seat belts reduces the incidence of accident fatalities.
 - D. CNG installations decline with the fall in the NZ dollar.

Fault(s) _____

[Return to index](#)

CAD Guidelines

3. Free market economic policies and the floating of the NZ dollar were designed to force commercial enterprises into being more competitive – analogous to Darwin's 'survival of the fittest' concept. In the short term, the effect of these policies has been to:
- A. Increase the value of the dollar and lower export potential.
 - B. Increase the value of the dollar and increase export potential.
 - C. Decrease the value of the dollar and lower export potential.
 - D. Decrease the value of the dollar and increase export potential.

Fault(s) _____

(This question appeared in a classroom vocabulary test prepared by a research student for 8-9 year olds.)

4. Which word is a synonym for adjacent?
- A. Behind.
 - B. Beside.
 - C. Underneath.
 - D. Above.

Fault(s) _____

5. Which statement is incorrect?
- A. Multiple choice questions are well-suited for the purpose of testing terms and essential facts.
 - B. Multiple choice questions decrease marking time in comparison with essay questions.
 - C. Multiple choice questions do not allow the testing of application skills.
 - D. Multiple choice questions increase guessing as a factor in student performance.

Fault(s) _____

(you will find the answers at the end of this guide)

Guideline 6 Provide plausible distractors.

Example: Poor question

Who originally posited the concept of the “Zone of Proximal Development” as an alternative to tests of knowledge recall?

- A. Anton Chekhov.
- B. Leonid Breshnev.
- C. Lev Vygotsky. *
- D. Alexander Luria.

If you write distractors that are implausible, you simply reduce the number of options from which the student need choose. In the above example, options A and B are too well known in their own fields to be plausible distractors.

Example: Better question

Who originally posited the concept of the “Zone of Proximal Development” as an alternative to tests of knowledge recall?

- A. Aleksey Leontyev.
- B. Ivan Pavlov.
- C. Lev Vygotsky. *
- D. Alexander Luria.

An effective strategy for creating distractors is to identify common errors of student thinking or application and to use these to form the basis of the alternatives provided. It can be helpful to create specific feedback statements associated with each incorrect answer that explain the likely source of the mistake and how students can learn to avoid the error in future. If you cannot create helpful feedback for an incorrect answer it is likely to be implausible.

If the questions are administered online this feedback process can be automated to provide students with more information and also supporting the use of questions for revision or self-tests. For example:

CAD Guidelines

Example: Distractors testing errors with feedback

Consider the following snippet of C++ code. What is printed out as a result of its execution?

```
int x=5;
for (y=0; y<3; y++)
{
    x--;
}
cout << x << "\n";
```

- A. 5
(feedback: The value of x is defined outside of the for loop and changes made within the scope of the loop are retained after it completes.)
- B. 2n
(feedback: The "\n" is a newline character equivalent to hitting the return key on your keyboard, the use of the '\' is called 'escaping' and normally indicates that a special or invisible character is being used.)
- C. 2
(feedback: Correct answer.)
- D. 1
(feedback: The loop executes three times with y=0, y=1 and y=2, after the third execution y is no longer less than the specified end value '3' and the code within the loop is not executed as the program continues with the output.)

If you have trouble thinking up plausible distractors, a useful tip is to use short answer questions instead, where students have to supply the correct or best answer. From the students' incorrect answers, you may be able to work out good distractors for a future multiple choice question.

Guideline 7 Don't give clues to the correct choice.

Example: Poor question

The type of test used to measure school achievement is called an:

- A. Diagnostic test.
- B. Ability test.
- C. Profile test.
- D. Achievement test. *

[Return to index](#)

CAD Guidelines

This question gives away an obvious clue by using the word 'achievement' in both the stem and option **D**. Another unintentional clue is the word 'an' at the end of the stem. This automatically eliminates options **A** and **C** which do not begin with vowels.

Example: Better question

The type of test used to measure school learning is called:

- A. A diagnostic test.
- B. An ability test.
- C. A profile test.
- D. An achievement test. *

The following tips help prevent three common mistakes which give clues to the correct answer.

1. See that options are all singular or all plural and that they match the stem in this respect.
2. Avoid using words such as "always" or "never". Students know that the world is seldom so black and white and will not select an option that contains such a word. Similarly, don't use words like "sometimes" or "often" because many events are 'sometimes' or 'often' true making such options likely candidates for the correct answer.
3. Don't make the correct option longer than the distractors in all your questions. This often happens when you add qualifying phrases to make sure the keyed option is really correct.

Examples of questions that illustrate these faults are given on pages 15 and 16.

Example of 1: Poor question

Which statistic would you use to indicate agreement between two variables?

- A. Correlation. *
- B. Measures of central tendency.
- C. Measures of dispersion.
- D. Cluster analysis.

Example of 1: Better question

Which class of statistical techniques would you use to indicate agreement between two variables?

- A. Correlation coefficients. *
- B. Measures of central tendency.
- C. Measures of dispersion.
- D. Cluster analyses.

Example of 2: Poor question

Which of the following is characteristic of standardised testing?

- A. It always involves comparisons with norms.
- B. It requires systematic procedures for the design and administration of tests. *
- C. It may involve the tailoring of test administration to fit with a student's age level.
- D. It necessarily involves consideration of the cultural background of students.

Example of 2: Better question

Which of the following is most characteristic of standardised testing?

- A. It involves comparisons with defined educational standards.
- B. It requires systematic procedures for the design and administration of tests. *
- C. It involves the tailoring of test administration to fit with a student's age level.
- D. It involves the creation of norms for different cultural groups.

CAD Guidelines

Example of 3: Poor question

The boiling point of water is:

- A. 212 C.
- B. 200 C.
- C. 120 C.
- D. 100 C at sea level in an open container. *

Example of 3: Better question

The boiling point of water at sea level, in an open container, is:

- A. 212 C.
- B. 200 C.
- C. 120 C.
- D. 100 C. *

Guideline 8 Don't use overlapping alternatives.

Example: Poor question

A major characteristic of a criterion referenced test is that it should:

- A. Discriminate between students.
- B. Determine readiness for learning.
- C. Spread students out.
- D. Identify the skills which a student has mastered. *

In this question, options **A** and **C** can be eliminated because they mean much the same thing. Be careful to see that options are clearly different and that one does not include or just give a more general description of another.

Example: Better question

A major characteristic of a criterion referenced test is that it should:

- A. Provide information for streaming students by ability bands.
- B. Determine readiness for learning.
- C. Spread students out.
- D. Identify the skills which a student has mastered. *

[Return to index](#)

Guideline 9 Don't ask students to express an opinion.

Example: Poor question

Which of the following people has contributed most to intelligence testing?

- A. Spearman.
- B. Thurstone.
- C. Guilford.
- D. Gardner.

Multiple choice questions are not suitable for measuring opinions because most options can be equally well defended. Essays are a much better way of measuring opinion because they let the student develop a line of argument.

Example: Better question

Which of the following people do you associate with the concept of "general intelligence"?

- A. Spearman. *
- B. Thurstone.
- C. Guilford.
- D. Gardner.

A related problem concerns the use of the word "you" in a question. In many contexts, "you" invites students to express an opinion even though a correct or best answer is available.

Example: Poor question

In which context would you use a taxonomy of psychomotor skills?

- A. In assessing inter-personal relations.
- B. In evaluating cognitive development.
- C. In assessing practical performance. *
- D. In judging the quality of test construction.

The use of "you" in this question implies that any answer is acceptable because "you" seems to be asking for the student's opinion only.

CAD Guidelines

Example: Better question

In which context would the use of a taxonomy of psychomotor skills be most appropriate?

- A. In assessing inter-personal relations.
- B. In evaluating cognitive development.
- C. In assessing practical performance. *
- D. In judging the quality of test construction.

Guideline 10 NEVER use 'all of the above'. Be careful when using 'none of the above'.

Example: Poor question

Which of the following terms is/are associated with the concept of assessing students' learning against pre-written statements of the knowledge or skills they should demonstrate?

- A. Outcomes based education.
- B. Criterion referenced assessment.
- C. Standards based assessment.
- D. All of the above. *

A student can answer this question on the basis of partial information. The student need only recognise two of the terms as correct to realise that the third must also be correct.

The problem is not overcome if only one option in a question is correct. Once a student identifies one incorrect option it then follows that 'all of the above' is also incorrect.

Example: Better question

Which of the following terms is associated with the concept of assessing students' learning against pre-written statements of the knowledge or skills they should demonstrate?

- A. Norm referenced assessment.
- B. Criterion referenced assessment. *
- C. Standardised assessment.
- D. External referenced assessment.

[Return to index](#)

CAD Guidelines

'None of the above' is more useful because students still have to consider each option before selecting this as the answer. However, do not use 'none of the above' in a best option format, as in the next example.

Example: Poor question

Which phrase is closest in meaning to the concept of reliability?

- A. Rank ordering of students.
- B. Correlation of test results.
- C. Consistency of test results. *
- D. None of the above.

In this question, the wording of the stem rules out 'none of the above' as an option. Even if all the other options were incorrect, one of them would still be closest in meaning to 'reliability'.

Example: Better question

Reliability refers to:

- A. The rank ordering of students.
- B. The sampling of course content and objectives.
- C. The consistency of test results. *
- D. None of the above.

One criticism of 'none of the above', even when it is properly used, is that incorrect selection of this option provides no diagnostic information to help you pinpoint what a student has misunderstood.

Guideline 11 Validate your questions after each use.

Validating your questions provides an opportunity to obtain feedback on their effectiveness and ensures that the same or similar questions can be used in the future with confidence.

The validation process looks at the choices students have made when answering a specific question in the context of their overall performance either in the test, or in the course as a whole. Provided that a sufficient sample of student responses are available (more than 20) it can identify issues needing some form of remediation.

For each question consider the proportion of students who have selected each possible answer. Compare this to the selection of answers by students who have received 'A' grades in the test or course. Consider whether the question passes each of the following tests:

- Is it answered correctly by students in proportions that seem consistent with its intended difficulty or challenge?
- Is it answered correctly more often by students with higher overall grades than by those with lower grades?
- Do students with higher grades choose predominantly only one of the possible answers?

Indications that a question may have problems include:

- The observation that students with higher overall grades fail to be differentiated from weaker students, suggesting that they have possibly misinterpreted an aspect of the question by looking beyond the 'obvious' interpretation.
- A bimodal distribution in answers (for all students as well as higher performing students) suggesting that potentially the question can be understood in more than one way.
- Very high proportions of students failing to choose the correct answer can mean that the question contains an error either in the main text or in the answers provided.

The information obtained from the validation process can be used to improve questions for future use and also can be a helpful guide in considering whether elements of the curriculum have been taught in ways that support learning by both stronger and weaker students.

Validation is essential if different versions of a test are used (with variations on questions, distractors, and/or the positioning of the correct answer) in order to discourage cheating. A common

CAD Guidelines

strategy is to alternate versions of a test when it is taken under supervision in a face-to-face setting. A large enough pool of validated questions can also be used to generate individualised online tests where each student answers a unique set of questions. This is helpful for reducing collusion and also can be used to generate practice or revision opportunities.

Guideline 12 Always have your questions checked by a colleague.

This simple precaution is likely to reveal the kinds of faults that are all too easily missed. Unseen ambiguities and unintentional clues to the right answer are often located at this step in the preparation of a multiple choice test. A colleague may also point out obvious omissions in your test content and suggest better ways of wording questions.

Additional tips

Tip 1: Use diagrams, graphs, tables, etc. where appropriate.

Some questions are best expressed in graphic or diagrammatic form because they are more understandable to students. Also, by giving information in this form you can often assess a student's ability to use problem solving skills (see Section 4).

Tip 2: Arrange the responses in a logical order.

Look at the following example. Which order do you find easier to scan?

Example:

The maximum temperature for frying chips is:

(random ordering)

- A. 220 F or 104 C.
- B. 400 F or 204 C.
- C. 360 F or 182 C. *
- D. 280 F or 113 C.

(logical ordering)

- A. 220 F or 104 C.
- B. 280 F or 113 C.
- C. 360 F or 182 C. *
- D. 400 F or 204 C.

CAD Guidelines

The numbers given in the right-hand column are shown in ascending order. Students can scan the options more quickly when they are arranged logically like this.

Tip 3: Study students' answers to short answer questions to identify plausible distractors for future multiple choice questions.

This tip has already been covered in the discussion of Guideline 6.

Tip 4: Vary where you put the correct option. Make sure that you don't set up a pattern of answers.

Evidence from research indicates that students in doubt tend to choose options **B** and **C**, so do not over-use these as the correct choice. In addition, an astute student may notice any pattern of correct answers. Have roughly equal numbers of correct responses in each option position, but distribute them randomly.

Tip 5: For long tests that are administered in a face-to-face environment, prepare a separate answer sheet and marking key.

As one of the major advantages of multiple choice questions is that they are quick to mark, it is sensible to go the whole way and prepare a separate answer sheet and scoring key. This is especially important if the test is long (e.g., 30 or more questions) or is sat by a large class of students (e.g., 50 or students).

►► ***Now do Exercise 3*** ◀◀
(starting on next page)

[Return to index](#)

CAD Guidelines

Exercise 3

Below you will see six multiple choice questions. Each question contains one or more of the faults discussed under guidelines 6 to 10 and tip 2.

Underneath each question, write down the fault(s) it contains. Use the following list.

- a. Implausible distractor(s) (guideline 6).
- b. Unintentional clue(s) (guideline 7).
- c. Overlapping options (guideline 8).
- d. Question asks for an opinion (guideline 9).
- e. 'All of the above' used (guideline 10).
- f. Inappropriate use of 'none of the above' (guideline 10).
- g. Responses not in a logical order (tip 2).

1. The port side of a ship refers to:
 - A. The left side of the ship when you are facing forward.
 - B. The side of the ship next to the wharf when the ship is in port.
 - C. The right side of the ship when you are facing forward.
 - D. The side of the ship where cargo is loaded from.

Fault(s) _____

2. A millisecond is:
 - A. 0.1 sec.
 - B. 0.001 sec.
 - C. 0.0000001 sec.
 - D. 0.01 sec.

Fault(s) _____

3. Who wrote the science fiction short story 'Nightfall'?
 - A. Ray Bradbury.
 - B. H.G. Wells.
 - C. Isaac Asimov.
 - D. Dylan Thomas.

Fault(s) _____

[Return to index](#)

CAD Guidelines

4. Which person has contributed most to our understanding of free market economics?

- A. Milton Friedman.
- B. J.M. Keynes.
- C. J.K. Galbraith.
- D. Ken Livingstone.

Fault(s) _____

5. Strikes most frequently occur in industries which:

- A. Have proportionately greater numbers of low paid workers.
- B. Have poorly organised union representation.
- C. Often control a vital resource.
- D. None of the above.

Fault(s) _____

6. For what purpose would you use a diagram to present information in a multiple choice question?

- D. To simplify the wording of the question.
- A. To test the ability of students to label correctly the parts of the diagram.
- B. To test the ability of students to interpret and apply information.
- C. All of the above.

Fault(s) _____

(you will find the answers at the end of this guide)

CAD Guidelines

STRENGTHS AND LIMITATIONS of multiple choice questions

Strengths

1. Multiple choice questions are quick to mark.
2. The scoring is accurate and objective.
3. Multiple choice questions are adaptable to most subject areas and levels of ability.
4. They allow you to test a wide coverage of course material because students have little writing to do.
5. They are better than essays for testing students whose verbal and language skills do not match their understanding of core skills and concepts.
6. Although multiple choice questions are used mostly to measure recall and comprehension, they can also be used to assess application and problem solving skills. See Section 4 for guidance.
7. Once written, questions can be stored for future use. Modifying an existing question or writing a parallel version of it, is much quicker than writing the original question. Thus, your work gets easier each time.
8. Multiple choice and other objective formats are useful for formative assessment. Sometimes a test given early in a course, or at an appropriate point in the course sequence, can highlight to students their misunderstandings. You can even arrange for students to self-check their answers as part of the learning process.
9. Questions can be analysed to see how difficult students found them, whether each question discriminates well between students on the basis of their overall test performance, and whether particular misunderstandings exist. The latter can be enhanced through analysis of the distractors that students select.

Limitations

1. Multiple choice questions are time-consuming to write. They are not justified for small classes.
2. Writing good multiple choice tests takes skill and experience. You must be thoroughly familiar with the subject content and well aware of all the traps you can fall into. Attention to detail is very important. You need to check your tests with reference to the above guidelines and tips.
3. People tend to write questions that test recall only; try to

[Return to index](#)

include questions that make your students think, not just remember facts.

4. Because options are provided, students will sometimes guess correctly answers which they do not know. Research indicates that this is only a minor problem if at least four options are given with each question. We recommend that you do not apply a correction for guessing unless you have a sound rationale for this decision.

3. Other objective question formats

A. True/false questions

These questions consist of a statement which students must judge to be true or false. Sometimes the options are right/wrong, yes/no, agree/disagree.

Example

T or F * Short answer questions are strictly objective in their marking.

True/false questions are useful for quickly testing a wide range of factual information such as vocabulary, technical terms and work and safety regulations. They are useful for formative assessment; students can self-check their responses against the correct answers. They are not recommended here for summative purposes unless you apply a guessing correction.

B. Multiple selection questions

These questions look like multiple choice questions. They consist of a statement and a set of options but with the important difference that more than one option may be correct. In effect, each option plus the stem is a true/false question.

Example

Which of the following are strictly objective questions? Circle all correct options.

- A. Multiple choice. *
- B. Essay.
- C. Short answer.
- D. Matching pairs. *

[Return to index](#)

CAD Guidelines

These questions are good for testing students in depth on a particular theme. They also save a great deal of repeated wording by grouping a number of true/false questions around a single theme.

In the above question, the student has to respond correctly to all four options. If they select both A and D, but not B and C, they have answered the question entirely correctly. Some writers suggest awarding 1 mark for a completely correct response (all four options responded to correctly), and ½ mark for making just one incorrect response (i.e., selecting the correct response for all but one option).

C. Assertion-reason questions

These questions consist of two true/false statements usually linked by the word 'because'. The first statement is an assertion and the second statement explains or justifies the first. The task of the student is to decide whether the first statement (the assertion) is true, whether the second statement (the reason) is true, and whether the reason explains the assertion.

Example

A take home test is suitable for testing recall of basic information because students can look up their textbooks for explanations.

(Answer: The assertion is false but the reason is true.)

These questions are good for testing whether students can see or reason out the relationship between different concepts or events. However, they are hard to write because they depend on natural connections between events.

D. Matching pairs questions

For these questions, students have to match each item in a list of words or phrases with an item from a second list.

CAD Guidelines

Example

Select from the right-hand column the description that best matches the kinds of testing given in the left-hand column. Write the matching letter next to each numbered item.

- | | |
|----------------------|--|
| 1. Closed book | A. Suitable for giving direction to student learning. |
| 2. Open book | B. Students are typically given 24 or 48 hours to return their answers. |
| 3. Issued-in-advance | C. Students must recall what they have learnt. |
| | D. Suitable for testing how quickly and accurately students can find information in manuals and regulations. |

(Answer: C, D, A)

These questions are ideal for matching technical terms with their definitions, diseases with their treatments, parts of diagrams with their labels, etc. They use testing time efficiently because they group together a number of related ideas under the one question. If you want to use this format for testing higher levels of learning, however, you have to put a lot of thought into preparing them.

You should also ensure that there are more options available (right hand column) than there are matches required for the list on the left hand column. If there were only three options in the question above, a student need only know the answer to two of the items to get the third right by elimination; a fourth option prevents this.

As in the discussion of multiple completion questions, some writers suggest awarding full marks for a completely correct response (all matches are correct), and part marks for making one incorrect match.

E. Sentence completion questions

In these questions, students complete sentences that have a word or phrase missing. To be strictly objective, you have to supply words or phrases to choose from, although you get few marking problems if you don't.

[Return to index](#)

CAD Guidelines

Example

Complete the following sentence. Select your answer from the words listed below.

A student answering a test consisting totally of _____ questions has about a 50% chance of getting half marks or better by guessing all answers.

- A. Multiple choice
- B. True/false *
- C. Assertion-reason
- D. Matching pairs
- E. Sentence completion

These questions are good for testing recall and understanding but are not very good for assessing application and problem-solving skills.

4. Writing objective questions to test more than recall

A. Write questions that test reasoning rather than memory.

Compare these two questions from physics:

The product, FORCE \times RADIUS, describes:

- A. Energy.
- B. Power.
- C. Torque.
- D. Work in rotation.

A given force is applied at right angles to the end of a spanner. The effective length of the spanner is 12 cm. A second force, half that of the first, is similarly applied to another spanner to produce a torque twice that of the first. What is the effective length of the second spanner?

- A. 3 cm
- B. 6 cm
- C. 12 cm
- D. 24 cm
- E. 48 cm *

[Return to index](#)

CAD Guidelines

The first question tests a student's knowledge of the formula for calculating torque. The second question extends this knowledge by asking the student to apply the formula to produce a result. However, the question goes a step further because it does not give values for the forces. The student must therefore reason out the answer based on the relationship 'half the forces produces twice the torque'. A very able student would seize upon this quickly, but most students would have to think through the information supplied very carefully before coming to an answer.

B. Write questions that test a student's ability to relate different concepts or to see the relationship between them.

A lot of your teaching is directed towards showing students the important links between the concepts or ideas in your subject. When you assess a relationship that you have taught, you are really requiring the student only to demonstrate recall or comprehension. Similarly, if you set tasks that involve routine application of a formula or method to data, even if the formula or method links different concepts, you are still measuring only recall and comprehension.

You are doing something quite different, however, if you test a relationship that you have not taught directly and you present it in a way that leaves the student to work out the link. Such questions assess a basic skill needed for problem solving – the ability to reason or analyse.

Here are two examples from statistics (example 1) and carpentry (example 2) that test whether a student can link ideas.

1. (Statistics)

T * or F: A Type 1 error is more likely to be committed when the significance level is set at .05 rather than .01.

Here students need to understand the definition of a Type 1 error and relate this to the concepts of significance and probability. If this relationship is not directly taught, students will need to think very hard about the connections involved.

[Return to index](#)

CAD Guidelines

2. (Carpentry)

A carpenter has available joists of 120 mm depth to do a job. A joist of 200 mm depth is needed to take the load spanning two piles. If all essential equipment is available, which of the following solutions will both do the job and support the greatest load?

- A. Bolt together two joists side by side.
- B. Glue laminate together two joists side by side.
- C. Bolt together two joists one above the other.
- D. Glue laminate together two joists one above the other. *

In this question, the student needs to work out, using the concept of 'load', the merits of joining joists side by side or one above the other, and the relative strengths of bolting and glue laminating the joists.

C. Introduce a novel or unfamiliar element into questions.

One way to encourage students to apply or transfer their learning to new situations is to set them questions that introduce a novel or unfamiliar element into the task. Such questions are useful for both in-class exercises as well as summative tests.

Here are two examples, one from office practice and the other from report writing. Questions such as these draw upon a student's ability to use judgment in a real situation. The second question also draws on a student's ability to analyse and classify information.

Example 1. (Office practice)

You have three urgent things to do, listed as A to C below. In what order should you do them? Write down the order in the space provided.

- A. Usher in to the boss a client who has been waiting five minutes.
- B. Ask a client who has been waiting ten minutes several details for your records.
- C. Acknowledge a new arrival whom you have not spoken to at all yet.

Write your answer here:

Answer: C, A, B _____

CAD Guidelines

Try the example, starting on this page, which tests your skill in relating numerical data to the guidelines given earlier in this guide on writing multiple choice questions. You will find that your answers will depend on your ability to analyse information and use reasoning to come to a correct response. You will find the answers given at the end of this guide.

Example: Item analysis

On the next page questions 1 to 4 refer to the table which shows the response patterns of students on three questions in a multiple choice tests.

Options for each question are listed as 'a, b, c, d'. The key is asterisked.

Students have been separated into two groups: the top 20 students on the test are classified as 'high scorers'; the bottom 20 students are classified as 'low scorers'.

For example, three of the high scoring students and six of the low scoring students selected option 'a' for question 10.

(Hint: If a question discriminates between the good and poor students, it should have more high scorers getting it right than low

CAD Guidelines

Question	Option	High scorers	Low scorers
10	a *	3	6
	b	2	5
	c	15	9
	d	0	0
17	a	0	1
	b *	18	17
	c	0	0
	d	2	2
30	a	8	4
	b *	9	5
	c	2	5
	d	1	6

Table 1. Response patterns of students on three questions in a multiple choice tests

- _____ 1. Which question is most likely keyed wrongly?
- _____ 2. Which question is most likely to have more than one correct answer?
- _____ 3. The explanation 'contains implausible distractors' best fits which question?
- _____ 4. For which question is inappropriate use of 'none of the above' most likely?

(Answers given on the last page of this guide)

5. Answers to exercises

Exercise 2

1.
 - i The 'not' in the stem should be emphasised.
 - d 'topic order diagram' is singular whereas 'They' in each option is plural.
 - c 'They allow tutors to' is unnecessarily repeated in each option.
2.
 - e The stem is too open and the options too diverse.
 - b Option A is unclear. Which age groups are being compared with 16 - 25?
3.
 - c Most of the first sentence of the stem is redundant.
4.
 - a 'synonym' and 'adjacent' are inappropriate for most 9 - 10 year olds.
5.
 - e The stem is too open.
 - i 'incorrect' is a negative and should be emphasised.
 - c 'Multiple choice questions' is unnecessarily repeated in each option.
 - g The 'not' in option C completes a double negative with 'incorrect' in the stem.
 - b Option D is unclear. What are multiple choice questions being compared with?

Exercise 3

1. c Options D and B overlap. Cargo is usually loaded from the side of the ship next to the wharf.
2. g Option D should appear second.
3. a Option D is clearly implausible.
4. d The answer to this question is a matter of opinion.
a Option D is implausible. Ken Livingstone is not a recognised economist.
5. b 'Often' in option C gives a clue to the right answer.
f 'None of the above' should not be used in a best choice format.
6. d The 'you' in the stem invites an opinion.
e NEVER use 'All of the above'.

Answers to the task for Writing Objective questions: 4.D pp 35—36

1. Question 10. More 'low scorers' than 'high scorers' picked option 'a'. Option 'c' appears to be the correct answer, with more 'high scorers' selecting this option than 'low scorers'.
2. Question 30. Both options 'a' and 'b' have response patterns that suggest they could be the correct answer. That is, in both cases, there were more 'high scorers' than 'low scorers' who selected these options.
3. Question 17. The very low response rates for options 'a', 'c', and 'd' indicate that each is likely to be implausible.
4. Question 10. No one has picked option 'd'. Alternatively, option 'd' could be implausible.