Good Practice Guide in Question and Test Design

Summary

These notes have been constructed as part of the PASS-IT Project for Preparing Assessments in Scotland using IT (http://www.pass-it.org.uk/). The first five sections draw heavily on the document entitled "Designing and Using Objective Tests" produced by Dr Joanna Bull and Dr Colleen McKenna as part of the Computer Assisted Assessment Centre (Teaching and Learning Technology Programme project, http://www.caacentre.ac.uk) at Luton University (1998 - 2001). However, the notes go beyond objective questioning based on the experiences of authoring non-objective tests at Heriot-Watt University during the CALM (http://www.calm.hw.ac.uk), Mathwise (http://www.bham.ac.uk/mathwise) and SCHOLAR Projects (http://scholar.hw.ac.uk) over the years 1985 - 2003. Sections 6 and 7 capture some of the CALM experience. Sections 8 - 10 deal specifically with questions taken from the PASS-IT Project itself in which the debate on question design is extended. So, the booklet is informed by almost two decades of computer aided assessment in a variety of subjects and should be viewed as complementary to the User Manual for the PASS-IT assessment system.

Introduction

This booklet will provide an introduction to good practice in question and test design. It includes the art of using objective tests to assess some aspects of student learning. At the outset you should be aware that objective tests are just one method of assessment. They are useful for assessing knowledge, comprehension and application and in some circumstances can be used to assess higher order skills such as evaluation and synthesis. They cannot be used to assess creativity, integration of ideas and the capacity to develop a sustained argument. Examples in this booklet range widely over a number of subjects and they include mathematical questions written as part of the Computer Aided Learning in Mathematics (CALM) Project at Heriot-Watt University in Edinburgh.

There is much evidence to show that assessment is the driving force behind student learning (Beevers et al (1991), Brown et al (1997)). It is increasingly the case that students are becoming ever more strategic in their study habits and they are reluctant to undertake work which does not count towards their final grade. Often, where additional material is made available to students for practice, it is the strong students who use the supplementary material and the weaker students who do not. Motivating students to undertake valuable but unassessed work is problematic. As assessment motivates students to learn, Computer Aided Assessment (CAA) offers the opportunity to test students more regularly without increasing the marking load on staff.

In this booklet, the following terms are used to classify methods of assessment:

- Diagnostic assessment --- tests which aim to determine a student's prior knowledge of a subject area or preparedness for a course of study;
- Self-assessment --- tests which show students whether they understand particular concepts and terminology where feedback is given but results are not recorded;
- Formative assessment --- an assessment for which the primary purpose is to promote learning by providing feedback, but which does not count towards the final mark or grade though marks may be recorded for information only or to bring the teacher into the process; and
- Summative assessment --- assessments which count towards the final mark or grade.

CAA lends itself well to a number of situations:

- Use of a multiple choice test to help students diagnose strengths and weaknesses used at the start of many university courses;
- Use of a range of question types to provide feedback to students as they work through some learning materials as in the Mathwise modules (see http://www.bham.ac.uk/mathwise);
- Use of a weekly test to provide informal feedback to teacher and student in continuous assessment as employed in the CALM Project (see http://www.calm.hw.ac.uk/); and
- A sequence of tests taken throughout a course of study which students must complete in order to qualify to take the module examination. This is the strategy being adopted by the Scottish Qualifications Authority (SQA) in their new Highers Still programme.

In practice, there may be areas of overlap between formative and summative assessment. For example, assessments taken during a course may contribute to the final grade but also provide feedback to students.

This booklet covers the following topics:

- The pedagogical issues associated with using objective testing;
- The construction of objective test questions, including examples of basic and complex question types;
- The scoring of objective tests;
- Using reports to analyse and improve questions and tests;
- Methods of organising questions into tests;
- The integration of objective tests with other assessments.
- Examples of simple mathematical questions;
- More complex mathematical questions involving random parameters; and
- Extended questions using steps to provide for partial credit.

Appendix A gives additional examples of objective test questions and Appendix B details relevant online resources.

Section 1: Why use objective tests?

1.1 What are objective tests?

Objective tests require a user to choose or provide a response to a question whose correct answer is pre-determined. Such a question might require a student to

- Select a solution from a set of choices (MCQ, true-false, matching, multiple response);
- Identify an object or position (graphical hotspot);
- Supply brief text responses (text input, word or phrase matching);
- Enter numeric text responses (number input); or
- Provide a mathematical formula (string evaluation or algebraic comparison).

Because the correct answers to objective test questions are pre-determined, they are well suited to the many forms of CAA that involve automated marking. The electronic marking of the responses is completely non-subjective because no judgment has to be made on the correctness or otherwise of an answer at the time of marking. However, it's worth remembering that in terms of in-built bias, an objective test is only as objective as the test's designer makes it.

Mathematical answers are non-objective in the sense that alternatively correct answers are permitted provided the answers are mathematically equivalent. For example, if x+1 is the answer to a particular problem and a student enters 1+x it would be a poor CAA system that did not mark the alternative equivalent answer as correct.

1.2 Background

Recent rises in student numbers have led to a re-evaluation of the type and frequency of assessments. Marking large numbers of student essays, projects and other subjective assessments is time-consuming, labour intensive and prone to errors of consistency (Newstead and Dennis 1994). Including objective tests within an assessment profile allows for more regular and efficient examining of students, because objective questions can be marked rapidly using computers, an OMR or staff members with no special knowledge in the area being assessed. Additionally, they eliminate the need for double (and triple) marking, thus saving time after the examination process. They also enable the lecturer to test a wide range of topics in a single assessment.

Objective testing is probably more widespread than most people realise. In the United States, for example, large-scale national objective tests, such as the American College Testing examination (ACT), the Scholastic Aptitude Tests (SAT), the Graduate Record Examinations (GRE) and the Law Society Admissions Test (LSAT) are used as performance indicators for all students seeking admission to undergraduate and postgraduate courses. These tests are regarded as the sector standard and have been in operation, in some cases, for over 40 years. They are increasingly delivered via computer networks in invigilated test centres across the country. (For more information, see the Education Testing Service website (http://www.ets.org).

In the United Kingdom, use of objective testing for formative and summative examinations is increasingly popular. The following list contains a small sample of disciplines and institutions, which employ objective testing, most of which involves computers or optical mark readers:

- Open University - large scale testing across a range of disciplines;
- UK medical schools - including University of Birmingham, University of North London, University of Leicester, University of Bristol, University of Dundee;
- Business and accounting – University of Ulster;
- Geology - University of Derby; and
- Mathematics - Brunel University, University of Portsmouth, Heriot-Watt University and Loughborough University.

Many universities now operate widespread objective testing in a range of departments including psychology, leisure, business economics, law, marketing, computing, information studies, physical and material science, biology, sports science, politics, engineering, chemistry, geography and accounting.

On a smaller scale, computerised objective testing is currently used in arts and humanities departments at Sheffield-Hallam, University of Sunderland, UCL and Kingston University.

Research by Don MacKenzie (University of Derby) in the TRIADS project is extending the capacity of objective tests by using sophisticated multimedia applications to construct question types, which are not feasible with traditional paper-based tests. It is claimed that many of the TRIADS questions also test higher learning levels (such as application, analysis, synthesis) and examples of TRIADS questions are used in this booklet. For more information about the project and available question types, visit the TRIADS website at http://www.derby.ac.uk/assess/talk/quicdemo.html.

1.3 Pedagogical issues

Objective tests are especially well suited to certain types of tasks. Because questions can be designed to be answered quickly, they allow lecturers to test students on a wide range of material. The use of CAA in the delivery of objective tests enables the provision of automatic feedback (in terms of scores, hints, praise, and guidance) to the student. Additionally, statistical analysis on the performance of individual students, cohorts and questions is possible.

The capacity of objective tests to assess a wide range of learning is often underestimated. Objective tests are very good at examining recall of facts, knowledge and application of terms, and questions that require short text or numerical responses. But a common worry is that objective tests cannot assess learning beyond basic comprehension. However, questions that are constructed imaginatively can challenge students and test higher learning levels. For example, students can be presented with case studies or a collection of data (such as a set of medical symptoms) and be asked to provide an analysis by answering a series of questions. If using a computer, students can be given electronic tools to manipulate or construct objects on a screen. Problem solving can also be assessed with the right type of questions. (See Appendix B.) Assessing higher learning levels with objective test questions will be considered more closely below.

A further worry is that objective tests result in inflated scores due to guessing. However, the effects of guessing can be eliminated through a combination of question design and scoring techniques. With the right number of questions and distracters, distortion through guessing becomes largely irrelevant. Alternatively, guessing can be encouraged and measured if this is thought to be a desirable skill. (See section 3 for a fuller discussion of designing of scoring issues.)

There are, however, limits to what objective tests can assess. They cannot, for example, test the competence to communicate, the skill of constructing arguments or the ability to offer original responses. Tests must be carefully constructed in order to avoid the decontextualisation of knowledge (Paxton 1998) and it is wise to use objective testing as only one of a variety of assessment methods within a module. However, in times of growing student numbers and decreasing resources, objective testing can offer a viable addition to the range of assessment types available to a teacher or lecturer.

Table 1.1 Advantages and limitations of objective tests:

The **advantages** to objective testing can include

- Significantly reduced marking time;
- Speed of assessment (A large number of questions can be asked quickly.);
- Wider coverage of topic content than essay questions; scope of test is broadened;
- Analysis of individual questions is possible;
- Provision of automatic feedback to student (when used in computer-based assessment);
- Potential for more frequent assessment; and
- Questions can be pre-tested in order to evaluate their effectiveness and level of difficulty.

The **disadvantages** to objective testing can include

- Significant amount of time required to construct good questions;
- Writing questions that test higher order skills requires much effort; and
- Cannot easily assess written expression or creativity.

1.4 Ways of using objective tests

There are a number of ways in which objective tests can be used in a module:

- Diagnostic --- Such tests may help a lecturer modify course content to fit the student's needs. And such tests help students and their teachers determine prior knowledge;

- Self-testing --- If mounted on a computer network, tests can be taken repeatedly and students can learn from their mistakes by the feedback provided. This is true formative testing.

- Continuous assessment --- Lecturers may wish to use tests at regular intervals within a course to determine which topics have been understood and to motivate students to keep pace with the teaching of the module. If the grades in such tests do not count towards some award then this form of testing remains formative. However, in order to encourage students to treat such assessments seriously, it could be a course requirement that in order to pass the module, a minimum score (40%, for example) would have to be obtained on all tests. Similarly, such tests could be used to encourage lecture attendance and regular learning.

- Summative --- Objective tests can be used to test the range of the student's understanding of course material. As above, it has been suggested that the capacity of objective tests to examine the breadth of a topic or module can be used to motivate good attendance because lecturers can assess a wider range of material in one examination (Kniveton 1996)). In the Scottish secondary educational system the Higher and Advanced Higher qualifications are obtained by passing a traditional examination at the end of the

course typically comprising three units. But, entry to the end of course examination has to be earned by passing minimum competency tests at the end of each unit. Such unit tests can be described as continuous since they occur over the period of the course but they are also summative since the grades count towards entry into the end of course examination.

1.5 What types of learning can be tested using objective tests?

Bloom's taxonomy of educational objectives is a useful starting point for categorising types of questions.

Table 2.2 Bloom's taxonomy and question categories

Competence	Skills demonstrated
Knowledge	• Recall of information • Knowledge of facts, dates, events, places • *Question words*: list, define, label, describe, name
Comprehension	• Interpretation of information in one's own words • Grasping meaning • *Question words*: interpret, discuss, predict, summarize, classify
Application	• Application of methods, theories, concepts to new situations • *Question words*: apply, demonstrate, show, relate
Analysis	• Identification of patterns • Recognition of components and their relationships • *Question words*: analyse, arrange, order, explain, connect, infer, compare, categorize
Synthesis	• Generalize from given knowledge • Use old ideas to create new ones • Organise and relate knowledge from several areas • Draw conclusions, predict • *Question words*: integrate, modify, invent, design, compose, plan, formulate, arrange
Evaluation	• Make judgements • Assess value of ideas, theories • Compare and discriminate between ideas • Evaluate data • *Question words*: appraise, judge, evaluate, defend, rank, conclude, discriminate, recommend

Based on the work by Benjamin B.S. Bloom et. al. *Evaluation to Improve Learning* (New York: McGraw-Hill, 1981)

It is commonly assumed that objective tests are only useful for examining the first three or four levels of learning. However, some educationalists, including McBeath (1992), suggest that all six levels can be tested using objective test questions. The questions on the following pages exemplify the different levels of Bloom's taxonomy.

Knowledge

Aim: to recall a specific date

Question: In which year did the American Civil War end?

a. 1832 c. 1865 *
b. 1857 d. 1888

Comprehension

Aim: to understand and interpret definitions, relationships and analogies

Question: In the following, a related pair of words is followed by five more pairs of words. Choose the response pair that best expresses a relationship similar to that expressed in the original pair.

ANTIDOTE : POISON

a. cure: recovery
b. narcotic: sleep
c. stimulant: relapse
d. tonic: lethargy *
e. resuscitation: breathing

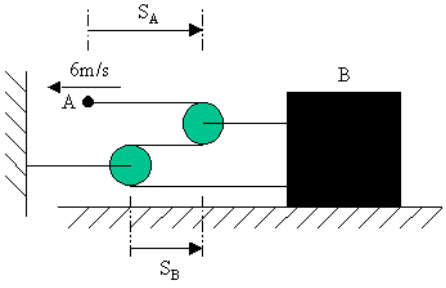
Above question from Graduate Record Examinations (GRE) - Practice Questions
http://www.ets.org/practice_test/gre/online.htm

Application

Aim: to calculate velocity

Question: End A of the cord is moving 6m/s to the left. Compute the velocity of the block B and select one of the responses below

4) End A of the cord is moving 6m/s to the left. Compute the velocity of the block B.

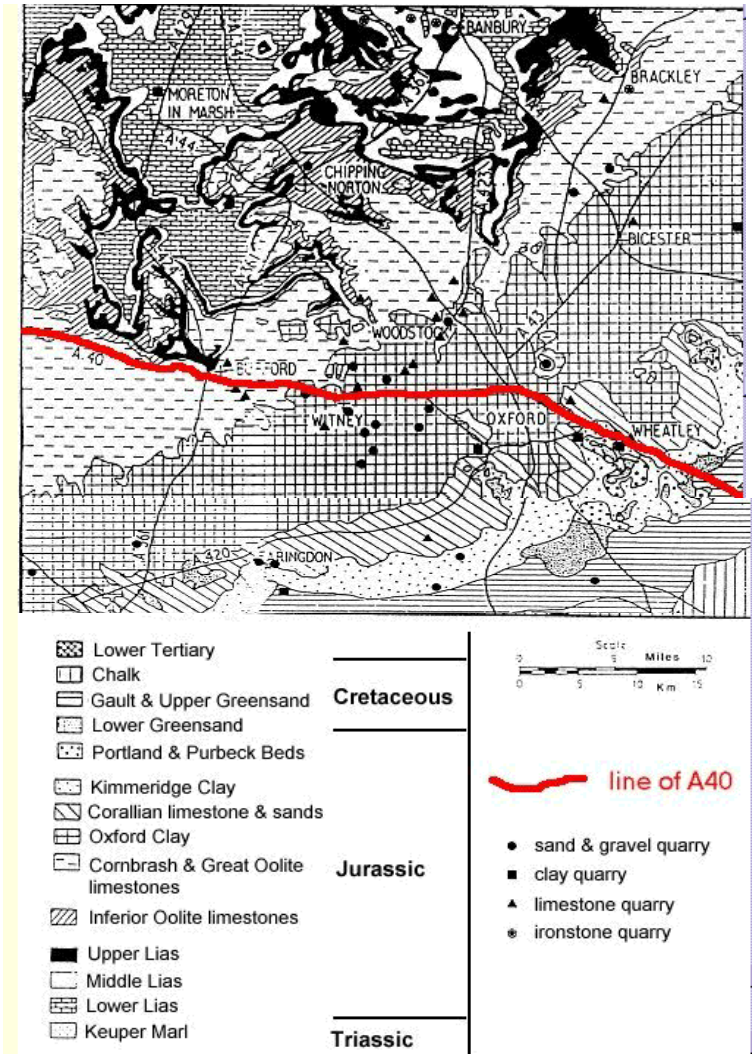


a. 2 m/s (to left)
b. 3 m/s (to left)
c. 6 m/s (to left) *
d. 12 m/s (to left)
e. None of the above

Question created by Rob Kay and Fiona Lamb at Loughborough University, Faculty of Engineering. (For a full range of their self-assessment MCQs in engineering, see the following site: http://www.lboro.ac.uk/faculty/eng/engtlsc/Eng_Mech/tutorials/tut_index.htm)

Analysis

Aim: to analyse and infer from a geological map



Study the geological sketch map of Oxfordshire (left), which shows the trace of the A40 road route, and read the notes on the geological problems encountered by engineers in building roads (linear routes). For the purpose of this exercise assume that the whole of this A40 route is going to be upgraded to motorway standards involving major reconstruction (in reality there is no current proposal to do this!).

These questions are part of a larger assessment created by Dr. Roger Suthren at Oxford Brookes University. The full assessment is located at http://www.brookes.ac.uk/geology/8307/a40frame.html

Question 1. LANDSLIPS:

At which of the following contacts between strata would you expect most landslips to occur?

Choose 2 of the following options.

a. Chalk above Gault Clay *
b. Gault Clay above Lower Greensand
c. Corallian limestone above Oxford Clay *
d. Cornbrash limestone above Inferior Oolite limestone

Question 2. SWELLING AND SHRINKAGE:
Which of the following rock units are most likely to show shrinkage and swelling with fluctuating dry and wet periods?

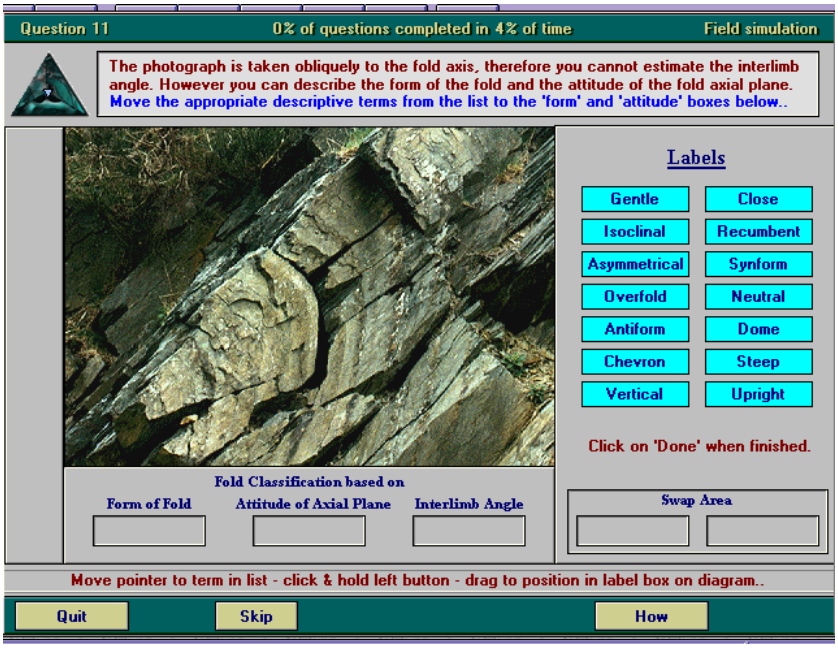
Choose 2 of the following options.

a. Chalk
b. Gault Clay *
c. Lower Greensand
d. Kimmeridge Clay *
e. Inferior Oolite limestone

Synthesis

Aim: to organise and arrange appropriate critical terms in order to construct a geological analysis of the following photographic image.

Question: Move the appropriate descriptive terms from the list to the 'form' and 'attitude' boxes below.



Question created by Professor Don MacKenzie, TRIADS. For a full demonstration of TRIADS assessment software go to http://www.derby.ac.uk/assess/talk/quicdemo.html.

Evaluation

Aim: to assess the condition of a patient based on various readings

Question: An adult subject breathing air was found to have the following lung volumes:

Vital capacity - 3 5 litres
Forced expiratory volume in 1 sec (FEV1) - 2.8 litres
Functional residual capacity (FRC) - 1.8 litres
Residual volume (RV) - 0.8 litres

True	False	
☐	☐	a. There is no obstruction to airflow.
☐	☐	b. The subject must be abnormal.
☐	☐	c. The expiratory reserve volume is 1 litre.
☐	☐	d. All of these measurements could have been made using only a spirometer
☐	☐	e. There would be approximately 250 ml of oxygen in this subject's lungs at the end of a tidal expiration.

Question by Dr David A. Davies, The University of Birmingham (MEDWEB Computer-Assisted Assessment web site - http://medweb.bham.ac.uk/http/caa/newdb/)

Section 2: Writing questions

2.1 Question types

Multiple-choice questions (MCQs) are the traditional 'choose one from a list' of possible answers.

True/False questions require a student to assess whether a statement is true or not.

Assertion-Reason questions combine elements of MCQ and true-false.

Multiple response questions (MRQs) are similar to MCQs, but involve the selection of more than one answer from a list.

Graphical hotspot questions involve selecting an area(s) of the screen, by moving a marker to the required position. Advanced types of hotspot questions include labelling and building questions.

Text/Numerical questions involve the input of text or numbers at the keyboard.

Mathematical questions can involve the input of an algebraic expression.

Matching questions involve linking items in one list to items in a second list.

Sore finger questions have been used in language teaching and computer programming, where one word, code or phrase is out of keeping with the rest of a passage. It could be presented as a 'hot spot' or text input type of question.

Ranking questions require the student to relate items in a column to one another and can be used to test the knowledge of sequences, order of events, level of gradation.

Sequencing questions require the student to position text or graphic objects in a given sequence. These are particularly good for testing methodology.

Field simulation questions offer simulations of real problems or exercises.

Other question types require students to identify and/or manipulate images. Students may be asked to plot a graph, complete a matrix, draw a line or build up an image using parts provided.

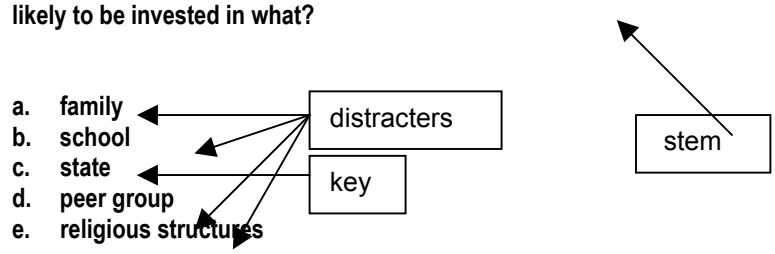
2.2 Multiple choice questions

A traditional multiple-choice question (or item) is one in which a student chooses one answer from a number of choices supplied. A multiple-choice question comprises:

- A **stem** - the text of the question;
- **Options** - the choices provided after the stem;
- The **key**: the correct answer in the list of options; and
- **Distracters**: the incorrect answers in the list of options.

Example of a multiple-choice question

As societies increase in complexity from folk to industrial, social control is more likely to be invested in what?



(*Question from Graduate Record Examination, Sociology Test 1997-1999*)

2.3 Suggestions for constructing multiple-choice questions

Writing stems

1. Present a single, definite statement to be completed or answered by one of the several given choices

Weak question: Genetically modified plants

a. Have been grown since 1985
b. Are free of organophosphates
c. Can cross with non-genetically modified plants
d. Require heavier rainfall

Improved question: Genetically modified plants require which of the following chemical enhancers?

a. Fertilizer
b. Pesticide
c. Soil improver

In the top example, there is no sense from the stem what the question is asking. The second example more clearly identifies the question and offers the student a set of homogenous choices.

2. Avoid unnecessary and irrelevant material

Weak question: Paul Muldoon, an Irish post-modern poet who uses experimental and playful language, uses which poetic genre in "Why Brownlee Left"?

a. Sonnet
b. Elegy
c. Narrative poem
d. Dramatic monologue
e. Haiku

Improved question: Paul Muldoon uses which poetic genre in "Why Brownlee Left"?

a. Sonnet
b. Elegy
c. Narrative poem
d. Dramatic monologue
e. Haiku

3. **Use clear, straightforward language in the stem of the item. Questions that are constructed using complex wording may become a test of reading comprehension rather than an assessment of whether the student knows the subject matter.**

Weak example: As the level of fertility approaches its nadir, what is the most likely ramification for the citizenry of a developing nation?

a. A decrease in the labour force participation rate of women
b. A dispersing effect on population concentration
c. A downward trend in the youth dependency ratio
d. A broader base in the population pyramid
e. An increased infant mortality rate

Improved question: A major decline in fertility in a developing nation is likely to produce which of the following?

a. A decrease in the labour forces participation rate of women
b. A dispersing effect on population concentration
c. A downward trend in the youth dependency ratio
d. A broader base in the population pyramid
e. An increased infant mortality rate

(Question from GRE - Sociology Test 1997-1999)

4. **Use negatives sparingly. If negatives must be used, capitalize, underscore embolden or otherwise highlight.**

Weak question: Which of the following is not a symptom of osteoporosis?

a. Decreased bone density
b. Frequent bone fractures
c. Raised body temperature
d. Lower back pain

Improved question: Which of the following is a symptom of osteoporosis?

a. Decreased bone density
b. Raised body temperature
c. Hair loss
d. Painful joints

5. Put as much of the question in the stem as possible, rather than duplicating material in each of the options. (Gronlund 1988)

Weak question: Theorists of pluralism have asserted which of the following?

a. The maintenance of democracy requires a large middle class.
b. The maintenance of democracy requires autonomous centres of countervailing power.
c. The maintenance of democracy requires the existence of a multiplicity of religious groups.
d. The maintenance of democracy requires a predominantly urban population.
e. The maintenance of democracy requires the separation of governmental powers.

Improved question: Theorists of pluralism have asserted that the maintenance of democracy requires which of the following?

a. A large middle class
b. Autonomous centres of countervailing power
c. The existence of a multiplicity of religious groups
d. A predominantly urban population
e. The separation of governmental powers

Question from GRE Sociology test book - 1997-1999.

Writing distracters

6. For single response MCQs, ensure that there is only one correct response.

Which of the following texts is considered to represent the pinnacle of modernist achievement?

a. The Waste Land
b. *Middlemarch*
c. "Ode to a Nightingale"
d. Ulysses
e. Ethan Frome

Improved: Which of the following texts represents one of the high points of modernist achievement?

a. The Waste Land
b. Middlemarch
c. "Ode to a Nightingale"
d. Ethan Frome
e. "My Last Duchess"

In the top example, both options **a** and **d** could be considered to be correct.

7. Use only plausible and attractive alternatives as distracters.

Weak question: Dichotic presentation of stimuli and shadowing are often used in the study of

a. Brightness constancy
b. Sensory scaling
c. Shadow dancing
d. Cartoons
e. Selection attention

Improved question: Dichotic presentation of stimuli and shadowing are often used in the study of what?

a. Brightness constancy
b. Sensory scaling
c. Illusions
d. Depth perception
e. Selection attention

Question from GRE-Psychology, 1997-1999.

8. Avoid giving clues to the correct answer.

A. Poor question: A fertile area in the desert in which the water table reaches the ground surface is called an

a. Mirage
b. Oasis
c. Water hole
d. Polder

B. Improved question: A fertile area in the desert in which the water table reaches the ground surface is called what?

a. Mirage
b. Oasis
c. Water hole
d. Polder

Example A uses the article “an” which identifies choice 2 as the correct response. Ending the stem with “what?” improves the question.

9. **If possible, avoid the choices "All of the above" and "None of the above". If you do include them, make sure that they appear as correct answers some of the time.**

   It is tempting to resort to these alternatives but their use can be flawed. To begin, they often appear as an alternative that is not the correct response. If you do use them, be sure that they constitute the correct answer part of the time. An "all of the above" alternative could be exploited by a test-wise students who will recognise it as the correct choice by identifying only two correct alternatives. Similarly, a student who can identify one wrong alternative can then also rule this response out. Clearly, the student's chance of guessing the correct answer improves as they employ these techniques. Although a similar process of elimination is not possible with "none of the above", it is the case that when this option is used as the correct answer, the question is only testing the students' ability to rule out wrong answers, and this does not guarantee that they know the correct one. (Gronlund 1988)

10. **Distracters based on common student errors or misconceptions are very effective.**

    One technique for compiling distracters is to ask students to respond to open-ended short answer questions, perhaps as formative assessments. Identify which incorrect responses appear most frequently and use them as distracters for a multiple-choice version of the question.

11. **Correct statements that do not answer the question are often strong distracters.**

    Avoid using ALWAYS and NEVER in the stem as testwise students are likely to rule such universal statements out of consideration.

12. **Do not create distracters that are so close to the correct answer that they may confuse students who really know the answer to the question. "Distracters should differ from the key in a substantial way, not just in some minor nuance of phrasing or emphasis." (Isaacs 1994)**

13. **Provide a sufficient number of distracters.**

    You will probably choose to use three, four or five alternatives in a multiple-choice question. Until recently, it was thought that three or four distracters were necessary for the item to be suitably difficult. However a 1987 study by Owen and Freeman suggests that three choices are sufficient (Brown 1997). Clearly the higher the number of distracters, the less likely it is for the correct answer to be chosen through guessing (providing all alternatives are of equal difficulty.)

2.4 Extending multiple-choice questions

Once you have tackled the common form of MCQs you may wish to try more complicated forms such as multiple true/false and reason assertion. Examples of these are given below.

Example of multiple true/false questions

In the following question the examiner is assessing whether the student can apply his/her knowledge:

A 45-year-old asthmatic woman who has lived all her life in Glasgow presents with a goitre of four years' duration and clinical features suggestive of hypothyroidism. Likely diagnoses include: -

a. Iodine deficiency
b. Dyshormonogenesis
c. Drug-induced goitre
d. Thyroid cancer
e. Auto immune thyroiditis

(Correct answer: true C and E: false A, B and D)

The student has to appreciate that in Great Britain iodine deficiency is not likely to be associated with hypothyroidism, that a 45 year old patient with only a four year history is unlikely to have dyshormonogenisis, that asthmatic patients not uncommonly take iodine containing preparations which may result in a goitre, that hypothyroidism is not usually associated with thyroid cancer and that auto immune thyroiditis typically is found in a middle aged woman with hypothyroidism.

(*From Brown, et al 1997.*)

Example of multiple true/false questions

In the following question the student's clinical judgement is assessed: -

A 28-year-old woman with one child has taken anti-thyroid drugs for 6 months for thyrotoxicosis. She has a friend who has been successfully treated with radioiodine. She finds she frequently forgets to take her drugs and wants to stop them to have radioiodine treatment.

a. She should be told that because of her age radioiodine is best avoided.
b. The problems associated with radioiodine should be discussed with her.
c. Surgery as a possible alternative should be discussed with her.
d. She should be advised that some form of further treatment is required.
e. You should find out more about her friend's treatment.

(Correct answer: true, B, C and D: false, A and E).

Examples from Harden, R.M. & Dunn, W.G. (1981) *Assessment a Work Manual Dundee Centre of Medical Education* in Brown *et al* 1997

Note: These approaches may be used for testing knowledge and judgement in many subjects.

When grouped together, a series of true/false questions on a specific topic or scenario can test a more complex understanding of an issue. They can be structured to lead a student through a logical pathway (Brown 1997) as in the above example, which simulates a medical diagnosis. Such questions may also be useful to the lecturer for diagnostic purposes, because they can reveal part of the thinking process employed by the student in order to solve the given problem.

Assertion-reason

The assertion-reason item combines elements of multiple choice and true/false question types. The method allows you to test more complicated issues and requires a higher level of learning.

The question consists of two statements, an assertion and a reason. The student must first determine whether each statement is true. If both are true, the student must next determine whether the reason correctly explains the assertion. There is one option for each possible outcome.

Example of assertion-reason question

Each question below consists of an assertion and a reason. Indicate your answer from the alternatives below by circling the appropriate letter.

	Assertion	Reason	
A	True	True	Reason is correct explanation.
B	True	True	Reason is NOT a correct explanation.
C	True	False	
D	False	True	
E	False	False	

Assertion

1. The blood sugar level falls rapidly after hepactectomy.
   BECAUSE
   The glycogen of the liver is the principal source of blood sugar.

2. Increased government spending increases inflation under all conditions.
   BECAUSE
   Government spending is not offset by any form of production.

3. Chloroform has a dipole moment
   BECAUSE
   The chloroform molecule is tetrahedral.

(Quoted in Brown *et al* 1997, p. 93 based on Matthews 1981.)

Assertion-reason tests can be used to explore cause and effect and identify relationships. When writing assertion-reason questions, keep in mind the following points:

- The reason should be a free-standing sentence so that it can be considered separately from the assertion;
- Avoid using minor reasons. These can result in an ambiguous question;
- Repeat options A-E in full for each question; and
- Use all five options as keys equally.

Multiple response questions

Multiple response questions are a variation of multiple choice in which the student is allowed to choose more than one choice.

Matching

Matching items require students to match a series of stems or premises to a response or principle. They consist of a set of directions, a column of statements and a column of responses.

Example 1 - matching test item

Directions: Column I contains descriptions of geographic characteristics of wind belts. For each statement find the appropriate wind belt in Column II. Record your answer in the appropriate space on the answer sheet. Answers may be used more than once.

Column I	**Column II**
1. Region of high pressure, calm, and light winds	A. Doldrums
2. The belt of calm air nearest the equator.	B. Horse latitudes
3. A wind belt in the northern hemisphere typified by a continual drying wind.	C. Polar easterlies
4. Most of the United States is found in this belt.	D. Prevailing easterlies
	E. Prevailing westerlies

Example from Ron J. McBeath ed. (1992) *Instructing and Evaluating Higher Education: A Guidebook for Planning Learning Outcomes* (New Jersey: ETP) 207.

Example 2 - matching test item

Directions: Match the quotation in column I with the literary school with which it is associated listed in column II. Items in column two may be used more than once.

Column I	**Column II**
1. You can lead a horse to water but you can't make it hold Its nose to the grindstone and hunt with the hounds. Every dog has a stitch in time. Two heads? You've been Sold One good turn. One good turn deserves a bird in the hand.	**A. Romanticism** **B. Modernism** **C. Neo-classicism** **D. Post-modernism** **E. Humanism** **F. Classical realism**
2. I cannot see what flowers are at my feet, Nor what soft incense hangs upon the boughs, But, in embalmed darkness, guess each sweet Wherewith the seasonable month endows The grass, the thicket, and the fruit-tree wile; White hawthorn, and the pastoral eglantine; Fast fading violets cover'd up in leaves; And mid-May's eldest child, The coming musk-rose, full of dewy wine, The murmur us haunt of flies on summer eves.	
3. frseeeeeeeefronnnng train somewhere whistling the strength those engines have in them like big giants and the water rolling all over and out of them all sides like the end of Loves old sweeeetsonnnng the poor men that have to be out all the night from their wives and families in those roasting engines stifling it was today Im glad I burned the half of those old Freemans and Photo Bits leaving things like that lying about he's getting very careless	
4. Twit twit twit Jug jug jug jug jug jug So rudely forc'd Tereu	
5. A perfect Judge will read each Work of Wit With the same Spirit that its Author writ, Survey the Whole, nor seek slight Faults to find, Where Nature moves, and Rapture warms the Mind;	

Advantages of matching questions

Matching questions are particularly good at assessing a student's understanding of relationships. They can test **recall** by requiring a student to match the following elements:

Definitions	-	terms
Historical events	-	dates
Achievements	-	people
Statements	-	postulates
Descriptions	-	principles (McBeath, 1992)

They can also assess a student's ability to **apply** knowledge by requiring a test-taker to match the following:

Examples	-	terms
Functions	-	parts
Classifications	-	structures
Applications	-	postulates
Problems	-	principles (McBeath, 1992)

Matching questions are really a variation of the multiple-choice format. If you find that you are writing MCQs that share the same answer choices, you may consider grouping the questions into a matching item. Tips for writing good matching questions include:

- Provide clear directions;
- Keep the information in each column as homogeneous as possible;
- Allow the responses to be used more than once;
- Arrange the list of responses systematically if possible (chronological, alphabetical, numerical); and
- Include more responses than stems to help prevent students using a process of elimination to answer the question.

True/False questions

True-false questions are a specialised form of the multiple-choice format in which there are only two possible alternatives. These questions can be used when the test-designer wishes to measure a student's ability to identify whether statements of fact are accurate or not.

Example of true/false question

> T F A poem with the following rhyme scheme could be correctly referred to as an English sonnet: abab cdcd efef gg.

True-false questions offer lecturers a very efficient method of testing a wide range of material in a short period of time. They can also be combined within a multiple-choice to create the more complex assertion-reason item. However, true-false questions do have a number of limitations:

- Guessing – a student has a 1 in 2 chance of guessing the correct answer of a question;
- It can be difficult to write a statement, which is unambiguously true or false – particularly for complex material; and
- The format does not discriminate among students of different abilities as well as other question types.

Suggestions for writing true-false questions:

- Include only one main idea in each item;
- As in multiple choice questions generally, use negatives sparingly;
- Try using it in combination with other material, such as graphs, maps, written material. This combination allows for the testing of more advanced learning outcomes (Gronlund 1988);
- Use statements that are unequivocally true or false;
- Avoid lifting statements directly from assigned reading, lecture notes or other course materials so that recall alone will not permit a correct answer; and
- Generally avoid the use of words that would signal the correct response to the test-wise student. Absolutes such as “none”, “never”, “always”, “all”, “impossible” tend to be false, while qualifiers such as “usually”, “generally”, “sometimes” “often” are likely to be true.

Text match response

The text match question requires a student to supply an answer to a question or complete a blank within a brief piece of text, using words, symbols or numbers.

Examples of text match question

a. Tony Blair is the leader of the _____ party.
b. 235 x 23 + (9+5) = _____.

A possible advantage of this question type is that the student must supply the correct answer rather than identify or choose it. The likelihood that the candidate will guess the correct answer is lower than that of a multiple-choice question. However, the short answer response questions can be difficult to phrase in such a way that only a single correct answer is possible. Additionally, if you are marking the assessments with computers, spelling errors may disadvantage students who know the right answer. (However, with some software, the test designer can identify numerous permutations of the correct answer for which the student will be awarded full marks.

For example if "United States" were the correct response to a short answer question, the author of the test could designate full marks to be awarded for all of the following: "United States", "US", "USA" and "United States of America".

Section 3: Designing responses and scoring strategies

3.1 Designing feedback for use with computerised assessment

One of the strengths of CAA is that students can receive swift feedback, which may offer some or all of the following:

- Scores (of individual questions and entire assessment);
- Notification of the correctness of a response; and
- Written guidance, praise and study suggestions based on a student's response.

Most CBA packages allow tutors to enter their own feedback, so that a message appears explaining to students why the response is wrong and perhaps directing them to relevant reading or course materials. Additionally, some packages allow for the inclusion of hints to be made available during the test to help a student toward the correct answer. (Marks can be adjusted accordingly.) In some cases, where students take summative examinations and receive no feedback, they have the option at a later stage to return to the examination and receive feedback at the end of each question.

The provision of feedback (particularly on wrong responses) is especially useful when using CAA for formative and self-assessments. The table below shows the alternatives for the following question and the feedback, which would appear beside the chosen response.

Table 3.1: Examples of feedback

Which of the following is the common plant name for the genus "Lonicera"?

Alternatives	**Feedback**
a. Laurel	No, the botanical name for the laurel is "Laurus".
b. Dogwood	No, the botanical name for the genus dogwood is "Cornus".
c. Honeysuckle	Correct.
d. Sunflower	No, the botanical name for the sunflower is "Helianthus".
e. London plane	No, the botanical name for the London plane is "Platanus".

3.2 Scoring objective test questions

Assigning scores to responses

A standard method of scoring responses would be to allocate one or more points for a correct response and no points for an incorrect one. Similarly, no points would be awarded for an unanswered question. You may wish to award more points per question according to the difficulty of the item.

Scoring complex question types

For more complex question types (such as multiple response or assertion reason) for which there may exist a range of correct or partially correct responses, scores can be assigned for different outcomes.

Examples of scoring a multiple response item.

The following question is a multiple response item. There are three correct responses to the question.

Which of the following are Irish poets? (Choose up to three responses.)

a. W B Yeats *
b. Carol Ann Duffy
c. Evan Boland *
d. Robert Frost
e. Tony Harrison
f. Seamus Heaney *

Score option 1:

Response a, c and f = 4 points.
All other combinations = 0 points.

Score option 2:

Response with all three correct answers a, c and f = 4 points
Response which includes two correct answers = 2 points
Response which includes any one correct response = 1 point
All others = 0 points

Score option 3:

A = +1
B = -1
C = +1
D = -1
E = -1
F = +1

This option corrects for guessing. Students should be advised that they will be penalised for choosing incorrect answers. (See “Negative marking” section below.)

For further general and technical information about scoring questions with multiple outcomes, see the Question Mark Authoring manual available at http://www.questionmark.com/perception/manuals/authoring_manual/index.htm.

Choose the link for "Constructing outcomes for multiple choice questions". Further discussion on this point can be found in the article by McCabe and Barrett entitled "The Mathematics of CAA" found at http://ltsn.mathstore.ac.uk/articles/maths-caa-series/apr2003/index.shtml.

3.3 Negative marking

A common anxiety about objective tests is that students will gain an inflated score simply through guessing. The effects of guessing can be largely eliminated through the use of corrective scoring in which marks are subtracted for incorrect answers. For example to discourage guessing, the responses to a multiple-choice question might be scored as follows:

+1 for the correct answer
-1 for an incorrect answer
0 for no answer.

Alternatively, questions could be scored as normal, and a formula for guess correction could be applied at the end. A standard formula is

SCORE = Right - Wrong/(n-1).

Right = the number of correct answers.
Wrong = the number of incorrect answers.
N = total number of alternatives per item (including the correct answer)

So, if a student answered 35 questions correctly on a 50-question test in which each item had 4 alternatives, the raw score after corrections would be calculated as follows:

Score = 35 - 15/(4-1)
Score = 30.

However, corrective (or negative) scoring is a contentious issue, and some experts feel that corrective scoring is unnecessary if the assessment has been well-constructed with good coverage of course content and a sufficient number of questions of varying difficulty. Alternatively, as Brown (1997) points out, intelligent guessing can be viewed as a useful skill in some subjects so lecturers in those subjects may wish to encourage it.

It is worth remembering that the relevance of guessing decreases as the number of test items increases. If you have a true/false test containing ONE question, a student has a 50% chance of getting it right and scoring full marks. The chances of scoring 100% on a 45- question test are less than one in a trillion (10^{12}) and the chances of earning 70% in a 45-question test are less than one in a thousand. (Heard et. al. 1997)

If you do decide to use corrective scoring, it is important to alert students to this, so that they can alter their test-taking strategies accordingly.

For a recent discussion of the perceived merits and disadvantages of corrective scoring, read the series of messages entitled "negative marks" located at http://www.mailbase.ac.uk/lists/computer-assisted-assessment/1999-02/thread.html in the computer-assisted-assessment mail base list.

Section 4: Analysing and improving objective test questions

One of the advantages of using objective tests is that a variety of analyses can be run to determine how well individual questions (and parts of questions) perform. This information enables the lecturer to improve or eliminate weak questions (such as those which are too easy or difficult.) Additionally, the performance of the test-takers (as a cohort and as individuals) can be analysed and this information can help the lecturer identify

- General problem areas within the module;
- Students who are struggling (see case study below); and
- Students who perhaps are not being sufficiently challenged.

While all of these analyses can be done by hand, some CAA software packages include a reporting facility that generates a range of statistical analyses on the performance of both the test-takers and the questions.

The following are the types of statistical measures, which help identify the usefulness of a particular question.

4.1 Discrimination

Discrimination helps to measure the power of an individual question to discriminate between the stronger and weaker students. It does so by looking at the correlation between the performance of a sub-group of students on a particular question and their overall scores on the assessment. There are a number of formulae for calculating discrimination. The following method (Brown *et al* 1997) looks at the number of correct responses to an item by the top and bottom thirds of the test-scorers (ranked according to their overall score on the assessment.)

D = (H-L)/N

D = discrimination
H = number of correct responses to question by the top third of test scorers
L = number of correct responses to question by bottom third of test scorers
N = number of students in the sub-group

If a test is taken by 90 students and 25 students in the top third answer the question correctly while only 10 students in the bottom third respond correctly, the discrimination of the question would be

D = (25-10)/30
D = 15/30
D=1/2 or 0.5

Discrimination scores range from -1 to +1. Positive scores of above +0.2 (a generally accepted lower bound) indicate a good discrimination and negative scores indicate a question with a poor discrimination. (Questions with a negative discrimination should be eliminated.)

4.2 Calculating the difficulty of questions using facility

Facility measures the difficulty of a question by dividing the number of correct responses by the total number of responses. Facility ranges from 0 to 1 (0 being very difficult with no one getting the answer correct and 1 meaning that all test-takers responded correctly) and can be calculated using the following formula (Brown *et al* 1997):

F = R/N

F = facility
R = the number of candidates who gave the correct response to the question
N = total number of candidates

Recommended values for F vary. Examples of acceptable parameters found in the literature are 0.2 to 0.9; 0.4 to 0.85 and 0.3 to 0.7.

Frequency analysis measures the number of times a choice or alternative has been selected. If certain distracters are rarely chosen, they may be ineffective and you should consider replacing them with more challenging alternatives.

4.3 Pre-testing questions

It is advisable to pre-test questions before they are incorporated into examinations and/or question banks.

A pre-test of the questions you wish to analyse would normally be taken by students who are representative of those who will be taking the actual assessment. Ask students to record their start and finish times. (If using CAA, this can be done automatically.) Following the pre-test, analyse the facility and discrimination of the individual items. You might also wish to analyse the frequency with which distracters within items are chosen and replace those which are selected infrequently, with more plausible options.

Analysis of the pre-test may also provide useful feedback on the module or course itself, by highlighting areas of difficulty.

One strategy for pre-testing would be to use questions one year as part of a formative assessment the score of which would not affect the student's module mark. After analysing and modifying the questions, they could be used the following year as part of a summative assessment.

If pre-testing is not possible, do ensure that the questions are free of ambiguity and that the instructions (particularly for more advanced question types, such as assertion-reason) are clear.

Beta-testing questions is another technique which can be used once you engage in a regular programme of objective testing. Include new questions in an assessment with known "good" questions. After the students have taken the assessment, evaluate the facility and discrimination of these questions. If these indicators suggest that the questions are good, then leave them as part of the assessment. If not, omit them from the calculations of scores.

4.4 Case study - sample analysis of student performance based on regular MCQ tests

The following example shows how the results of frequent objective testing can be used to identify possible problems with student performance:

The table below shows an extract from a spreadsheet recording the results of a group of sociology students in a weekly multiple-choice test designed to reinforce lecture material.

Table 4.1 Week no. and assessment scores (percentage)

Students	1	2	3	4	5	6	7
A	57	63	21	35	40	27	20
B	68	71	45	79	83	80	77
C	23	21	11	0	0	0	0
D	45	51	45	79	83	80	77
E	0	0	0	0	0	0	0
F	63	0	51	0	47	0	35
G	54	58	35	50	58	60	62

The pass mark is 40%.

Sample analysis of the results:

The table provides a great deal of additional information about students A-G over the seven-week period.

Student A's performance has diminished steadily over the seven-week period and is a cause for concern. Most people would share the view that Student B is one we have little to worry about. Student C, however, has failed the first three tests and after Week 4 has failed to undertake any assignments at all. This is a student we would need to look for and investigate. Student D was not doing well to start with but after Week 4 is achieving results identical to those of Student B. This is indeed a cause for concern and suggests either a case of advanced telepathy or cheating! Most tutors would recognise Student E as one who causes a great deal of work. Is this student on the course or did he or she drop out at an early stage? Student F obviously has a problem. It seems likely that this student is either doing a part-time job on alternate weeks or is suffering from the 'cheap British Rail fares on Thursdays' syndrome and is therefore missing Friday afternoon classes. At the same time this student's test scores are diminishing rapidly so she or he needs a great deal of attention. After a glitch in week three, student G is making steady progress.

Example from *Strategies for Diversifying Assessment in Higher Education,* Oxford Centre for Staff and Learning Development, Oxford Brookes University, 1994

Section 5: Constructing tests

5.1 Matching assessment methods to learning outcomes

Before writing your questions, analyse your assessment needs and determine WHAT you wish to assess and HOW you wish to assess it. Identify the learning outcomes of your module and determine which components can be assessed using objective tests. Consider integrating your objective assessment with other methods so that you cover an entire range of learning levels and competences

Prior to organising your assessment, briefly analyse your question material according to course content, difficulty, learning level, assessment type and question style. Such a breakdown of assessment components can help you identify potential areas of low coverage and unequal spread – either in terms of content or question type.

Additionally, consider the student profile for the module:

- Will the students come from varying academic backgrounds?
- Will some need more practice in certain areas than others?
- Do you need to know what their knowledge base is before the module begins?
- Would it be useful to receive feedback as the term progresses on the extent to which students understand the material?

If the answer were yes to any of the above, then using objective tests for diagnostic and formative assessments would be useful.

Below are examples of the types of tables that can be used for analysis.

The following are assessment breakdowns for Introduction to Poetry, a 12-week, first year undergraduate module. The lecturer wishes to assess the students in the following areas:

- Knowledge and application of critical terms and theories across a range of topics;
- Awareness of the process of development of a poem from manuscript form to publication;
- Ability to write a critical comparison of two poems; and
- General understanding of the historical development of poetry.

The lecturer leading the module has to plan the assessments with the following issues in mind:

- The cohort will be large (approximately 140 students);
- The previous experience of the students will be mixed; some will not have formally studied poetry before;
- For some, the module is an elective: they do not intend to take their study of literature further; and
- For others, the module is a pre-requisite for second and third-year literature modules.

Table 5.1 Module – An Introduction to Poetry (year one)

Overall assessment plan

Topics for assessment	**Assessment methods** (✓ indicates the suitability of method for each topic.)			
	Objective testing	Critical comparison of two poems	Essay response	CAL manuscript study assignment
Knowledge of poetic terms associated with register, imagery, rhythm, and form	✓✓✓ (Lecturer can ensure students answer questions on all terms.)	✓✓ (Some, but not all, terms will arise in the discussion of two poems.)	✓✓ (Some, but not all, terms will arise in essay discussion.)	✓ (Will entail fairly limited discussion of terms.)
Application of poetic terms and concepts	✓✓✓ as above	✓✓ as above	✓✓ as above	✓ as above
Understanding of historical development of genre	✓✓	✓	✓✓✓	✓
Ability to critically compare two poems	✓	✓✓✓	✓✓	NA
Understanding of poetic process from manuscript to publication	✓	NA	✓✓	✓✓✓

The objective tests for the module may be formative, diagnostic and summative. A sample breakdown of an objective assessment is in the table 5.2 below.

Table 5.2 Sample objective assessment on imagery – 35 questions

	Bloom Learning Levels – no. of questions					
Topic	Knowledge	Compre-hension	Application	Analysis	Synthesis	Evaluation
Metaphor	1	2	2	2		
Symbol	1	2	2			
Personification	1	1	2	2		
Synecdoche	1	1	1	1		
Defamiliarisation	1	2	1	2		
Genre - haiku, imagist poem, etc.	1	2	2	2		

Comments: The above assessment on imagery could be taken by the students after a two-week session on the topic. By analysing the results of the test, the lecturer could determine whether students understand the relevant terms and concepts. Furthermore, such an assessment requires all students to answer questions on all topics. If objective testing were eliminated from the above assessment profile, it would be possible for students to avoid being assessed on entire sections of the module through tactical selection of essay questions.

5.2 Timing

One of the potential strengths of objective tests is that they enable a relatively large number of questions to be asked in one assessment, thereby increasing the coverage of topics. However, when writing a test, the designer should be careful to allow enough time for the well-prepared student to complete the test, particularly if it includes complex questions which require a lengthy reading time. One way of resolving this problem is to present a case study about which several questions are asked.

As a general rule, the time allowed for an objective test should not exceed 90 minutes. If you are examining higher levels of learning than recall, then plan for no more than 40 items per hour. Pre-testing the questions will help you gauge the time needed to complete them. It seems obvious to note that performing similar questions on paper and on computer will not take the same time. In fact, it is generally true that students take longer to do questions on computer than equivalent questions on paper.

5.3 Question order

Grouping like items

When assembling the test, you may wish to group similar items together. Items can be grouped according to

- Question type (MCQ, T/F, short answer, etc.);
- Measurement of learning level (knowledge, comprehension, application, analysis, etc.); or
- Subject matter.

It is also advisable to arrange questions in order of increasing difficulty, in order to give students confidence at the early stages of the section and to prevent weaker students from spending too much time on difficult items.

Test-item sequencing

There is conflicting evidence about the extent to which the sequencing of questions relative to the order in which the material was taught can influence the student's performance on the test. Baldwin and Howard (1983) found that accountancy students who sat an examination in which the questions were arranged in the same order in which the material had been covered in the module performed better than students taking the same questions arranged in a random order. By contrast, Gruber (1987) argues for a "recency effect" to explain findings of a study in which students who sat tests in which questions had been arranged in reverse order in which they were presented in the module (with the most recent coming first) performed better that those who took the same test with the questions ordered randomly or in the same sequence in which they were taught. Stout and Wygal (1990) found that test-item sequencing relative to the order in which topics were taught had little effect on the performance of accountancy students.

Randomising test-items with question banks

When administering summative examinations, it may be desirable to randomise questions in order to reduce the likelihood of cheating between students who are sitting in close proximity. Some CBA software allows for the random generation of test questions. However, if the scores are to be used for grading then care should be exercised since the student experience is not truly a level playing field. Randomisation in formative assessment is more acceptable.

You may also wish to use random selection in combination with question banks, which are collections of questions normally on one subject. In some software packages, items can be chosen at random from question banks, thus allowing a unique subset of questions to be chosen for each assessment or student. This is particularly useful for formative testing and enables students to test themselves on the same topic repeatedly.

Questions can also be grouped into banks according to difficulty or the type of skill being tested, such as recall, comprehension, analysis or application. Assessments can then be designed which draw a certain number of questions from each bank, thereby ensuring that, while the test is randomised, specific skills and levels of competence will be examined.

It is useful to use banks in accordance with your test specification matrix.

Section 6: Mathematical Questions

As mentioned in the summary this booklet now seeks to gather together some of the experiences of mathematical testing as practised in the CALM Project at Heriot-Watt University since 1985. Specifically, in the next two sections there will be descriptions of the use of mathematical expressions as answers, the place of steps to provide scaffolding in formative assessment and as a source of partial credit in summative assessment and the role and use of random parameters will also be explored. Before PASS-IT began the internal Heriot-Watt assessment system was known as CUE, to reflect the collaboration between the CALM group, the University of Cambridge Local Examination Syndicate and the commercial company EQL International of Livingston in Scotland. It should be noted that EQL International have their own commercial version of this assessment system known as i-Assess and details can be found at http://www.i-assess.co.uk).

The Mathwise assessment engine which ran on PC only contained a range of desirable features, including randomised questions, flexible feedback, algebraic marking and a mathematical input tool. To our knowledge no other computer-based system had incorporated such an extensive and sophisticated set of tools by the middle 1990s. However, the development work needed to keep pace with computer advances was not possible in the distributed Mathwise Project. The necessary development work continued within the CALM Project and the latest version of the assessment engine has an easy to operate editor for question construction. Tests can be delivered over the web using an NT, Windows 2000 or XP server.

One further educational feature in CUE, Mathwise and i-Assess that is worthy of note is that the tests can appear in a variety of modes. In Examination mode there is no visible marking on the screen though the automatic assessment continues in the background. But, the same test can be set in Help mode in which ticks and crosses appear immediately at the end of each step or part and the student has the chance to reveal an answer if they are really stuck. This feature has again been very popular with the students particularly those who are independent learners.

6.1 A simple mathematical question

In this example consider the question which asks the student to write down the equation of a tangent to a curve at a given point.

Question 6.1: Find the tangent to the curve given by $f(x) = x^3 + 4$ at the point where $x = 1$.

Key Part (KP)1. Equation of tangent is y = ? 3x + 2 (1)

The first thing to notice here is that the student has to type in the response as an algebraic expression, which is then compared with the correct answer stored on the computer. The method of string evaluation has been adopted by the CALM group at Heriot-Watt University and the interested reader can find a discussion on this and related matters in the references: Beevers et al (1991) and Beevers et al (1992). The important point about string evaluation is that any mathematically equivalent correct answer will be rewarded with a tick. So, the question setter does not have to put down all possible correct answers, as would be the case for string recognition or text match techniques. The de facto standard one-line input favoured by graphics calculators has been adopted in the work of the CALM group since 1985, as students are familiar with that format. In addition, projects like Mathwise (see http://www.bham.ac.uk/ mathwise/) and Interactive PastPapers (details are also on http://www.calm.hw.ac.uk) have followed a similar approach. One concession to computer notation has been made to make the one-line input consistent with the usual written mathematical convention. This concession refers to implied

multiplication which is permitted input ensuring for example that *3x* and *3*x* are entirely equivalent.

6.2 Questions with random parameters

As indicated in section 6.1 it is possible to check for mathematical answers by using string evaluation techniques. It is also possible to include numerical randoms in Maths display and answers. The following example in multiplication of complex numbers illustrates the role of randoms in mathematical CAA.

Question 6.2: Find the complex number z where

z = ({a} + {b}i)({c} + {d}i)

expressing the answer in the form x + yi with x and y real.

KP1. z = ? {a*c-b*d} + {b*c+a*d}i (1)

Note that the use of the curly brackets {a} etc indicates a random parameter which can be chosen to take, for example, the values -3, -2, -1,1,2,3 randomly and so would appear differently each time the question is run. The answer would also appear as numbers and not letters taking on the value indicated by {a*c-b*d} etc. So, for example with a = 1, b = -2, c = 3 and d = -1 then the answer would be calculated by forming z = (1-2i)(3-i) = 1 - 7i giving the instance of the example above as:

Question 6.3: Find the complex number z where

z = (1 - 2i)(3 - i)

expressing the answer in the form x + yi with x and y real.

KP1. z = ? 1 - 7i (1)

For checking purposes answers with i can be treated as a function of variable i and the string evaluation technique again works well. The use of random parameters, particularly in Science and Engineering subjects, provides students with plenty of practice since each time they take the question the numbers have changed.

6.3 Multi-stage mathematical question

In another sophistication adopted by the CALM Project team some questions take the user through a mathematical question in stages. The student user is prompted to an answer through a series of key parts.

Following the examples in the previous parts of this section it is possible to set more complicated mathematical questions which ask for more than one answer. For example:

Question 6.4:Find the equation of the tangent to the curve defined parametrically by

$$x = \{a\}t^3 + \{b\},\ y = \{3c\}t^2 - \{d\}$$

at the point where t = 1. Write down a formula for $\frac{dy}{dx}$ in terms of t and construct the equation of the tangent in the form y = mx + c at t = 1. Also, compute the value of $\frac{d^2y}{dx^2}$ at the point t = - 1.

KP1. dy/dx (in terms of t) is ? {2c}/({a}t) (1)

KP2. Equation of tangent at t = 1 is y = ? {2c}x/{a} + {c - d - 2c*b/a} (1)

KP3. Value of d^2y/dx^2 at t = -1 is ? *{- 2c/(3a²)}* (1)

This question has been set with four random parameters and for ease of explanation here let us suppose that

a = 1,3,5;
b = 2,4,7,8;
c = 2,4,8;
d = 1,3,5,7

so each time this question is set different numbers appear on the screen. Other combinations of the randoms are possible within the PASS-IT assessment system and it can provide simplifications to reduce, for example, fractions to their neatest form. This question is typical of the questions set for Engineering and Science undergraduates in a course on parametric differentiation.

6.4 Question design using optional steps

Assessment is an essential part of learning and students progress at different rates. Another educational feature exploited by the CALM team involves the introduction of intermediate optional steps to help students answer the key parts of a question. The optional steps can help students accrue marks with the combined mark being allocated to the key part if the student chooses to work only through the key parts.

This therefore performs 2 functions, that of breaking down the question for those who are unsure of how to proceed and of enabling partial credit to be awarded. Please note that in summative assessment the problem of partial credit is significant (Beevers et al (1995)). A student may not be able to answer a question in full, but there is a case for some marks being awarded for a partially correct answer. This point is addressed in the CALM Project, Mathwise and i-assess by allowing the student to do a question through a series of key parts and optional steps. This means that the student may opt to break a lengthy question down into sub-parts or steps and

then attempt to gain partial credit in each part answered. In a later paper Beevers et al (1999) considered other forms of partial credit available in automatic testing of mathematical ability.

As an example consider a small amendment to the specific example on complex numbers shown as example 6.2 above.

Question 6.5: Find the complex number z where

z = (1 - 2i)(3 - i)

expressing the answer in the form x + yi with x and y real.

KP1. z = ? (2)

<STEPS>

The only obvious difference between this version of the question and its previous form is the appearance of the **<STEPS>** button. Some students may have forgotten the basics of complex algebra and the prompt which appears at the press of the **<STEPS>** button may well remind them how to get started. So, on pressing the **<STEPS>** button the optional step is revealed as follows:

Question 6.5: Find the complex number z where

z = (1 - 2i)(3 - i)

expressing the answer in the form x + yi with x and y real.

Step1.1. What is i*i ? *–1* (0.5)

KP1. z = ? *1 - 7i* (1.5)

With 0.5 marks for the step and 1.5 for the key part this question is worth 2 marks. The student who chooses to answer this question via the key part only can do so and still gain the full two marks. In fact the insertion of the **<STEPS>** button does not impede the good student's progress to the correct answer but it does allow a weaker student the chance to receive a hint or prompt to help them get started.

Consider next a randomised version of question 6.1 about tangents which appeared earlier in section 6.1.

Question 6.6: Find the tangent to the curve given by

$f(x) = x^{\{a\}} + \{b\}$

at the point where x = 1.

KP1. Equation of tangent is y = ? *{a}x + {b+1-a}* (2)

<STEPS>

The student is faced here with a single key part question worth say two marks. However, the **<STEPS>** button allows a choice to be made if the student cannot answer this question through the one key part. When the **<STEPS>** button is chosen then the question appears with three extra but smaller steps before key part 1 as follows:

> **Question 6.6**: Find the tangent to the curve given by
>
> $f(x) = x^{\{a\}} + \{b\}$
>
> at the point where x = 1.
>
> Step 1.1. Derivative of f with respect to x is ? $\{a\}x^{\{a-1\}}$ (0.25)
> Step 1.2. Gradient at x = 1 is ? *{a}* (0.5)
> Step 1.3. y-coordinate at x=1 is ? *{b+1}* (0.25)
>
> KP1. Equation of tangent is y = ? *{a}x + {b+1-a}* (1)

If the student selects the **<STEPS>** button then there are three further tasks before key part 1 and this might help jog the memory. With steps 1.1 and 1.3 carrying say 0.25 marks each, step 1.2 assigned 0.5 marks and key part 1 a further 1 mark then the student who answers the question through the single key part would be awarded two marks but fractions of the total mark are also available by this more flexible approach. In formative assessment this allows the students more help as they attempt the question, and in summative assessment such additional parts provide the student with a chance to score some partial credit. In our experience this flexibility in formative assessment allows more students to progress. In summative assessment it introduces another examination technique which students have to consider: if they choose to take the steps then they must enter those extra answers since looking at the steps but not answering them would cost the student one mark. Hence, students have to balance time with the chance to score some partial credit.

It should be noted that in the original curve set in Question 6.1 with the function $f(x) = x^3 + 4$ then {a} = 3 and {b} = 4 but in this randomised version both {a} and {b} could take the values 2..9 (the integer values 2 to 9) randomly each time the question is set. It is also worthy of note that the answer *{a}x +{b+1-a}* could have been entered as *{b+1-a} + {a}x* or *{a}(x - 1) + {b+1}* as mathematically equivalent correct answers. In addition, the PASS-IT assessment engine has a range of options available to the question setter to ensure that correct answers take a prescribed form, fraction not decimal for example. In the above example the question setter who wants to prevent the untidy answer *{a}(x - 1) + {b+1}* could insist that brackets are not allowed or limit the answer to a fixed number of characters. For more details on this sophistication the interested reader is directed to the manual that accompanies this guide.

As a further illustration of key parts and optional steps consider the example first set out in section 6.3 in key parts only. Here some optional steps have been added and are all displayed as an alternative approach to answering this question through three key parts only:

> **Question 6.7**: Find the equation of the tangent to the curve
>
> $x = \{a\}t^3 + \{b\}$, $y = \{3c\}t^2 - \{d\}$
>
> at the point where t = 1.

Write down a formula for dy/dx in terms of t and construct the equation of the tangent in the form y = mx + c at t = 1. Also, compute the value of d^2y/dx^2 at the point t = - 1.

Step 1.1. dx/dt = ? *{3a}t*^{2} (0.5)
Step 1.2. dy/dt = ? *{6c}t* (0.5)

KP1. dy/dx (in terms of t) is ? *{2c}/({a}t)*b (1)

Step 2.1. x coordinate at t = 1 is ? *{a + b}* (0.25)
Step 2.2. y coordinate at t = 1 is ? *{3c - d}* (0.25)
Step 2.3. Value of dy/dx for the curve at t = 1 is ? *{2c}/{a}* (0.5)

KP2. Equation of tangent at t = 1 is y = ? *{2c}x/{a} + {c - d - 2c*b/a}* (1)

KP3. Value of d^2y/dx^2 at t = -1 is ? - *{2c}/{3a*^{2}*}* (2)

Key parts 1, 2 and 3 would each carry 2 marks if a student chooses to answer this question via key parts only giving a total of 6 marks for the whole question.

However, using steps 1.1, 1.2 and 2.3 each worth 0.5 Marks and steps 2.1 and 2.2 worth 0.25 marks each then key part 1 and 2 have one mark for them with key part 3 worth 2 marks. Hence there are again a total of 6 marks but a student may choose to gather the marks more slowly through these optional steps. This approach has been welcomed by the students in both formative and summative assessment. In the former extra scaffolding in some questions does help to consolidate the learning for some students whereas in summative assessment this optional approach enables students to score some marks even if they cannot complete the whole question. In formative assessment it is possible to add optional steps that carry no marks and then the user can treat the steps as if they are hints.

6.5 Further adaptations using optional steps

In this final part of Section 6 an example of percentage profit is chosen to illustrate alternative approaches to designing questions in a summative test.

Suppose that the learning outcome in a course is to be able to determine the percentage profit on a sale. Consider the example:

Question 6.8: A greengrocer buys a bag of nine oranges for £1.50 and sells them on at 29 pence each. What is the percentage profit on the sale when all nine oranges have been sold?

KP1. Percentage profit is ? *74* (2)

As explained in earlier sections this version of the question is fine for a good student who can make the intermediate calculations needed to complete the answer and gain the two marks on offer but it does not help the average student during formative assessment exercises. Nor does it provide any opportunity for partial credit in a summative test. An alternative version of the question can be created with optional steps and would appear as follows after pressing the **<STEPS>** button:

Question 6.9: A greengrocer buys a bag of nine oranges for £1.50 and sells them on at 29 pence each. What is the percentage profit on the sale when all nine oranges have been sold?

Step 1.1. How much in pence is there when all nine oranges are sold? *261* (0.5)
Step 1.2. What is the profit in pence when all nine oranges have been sold?*111* (0.5)

KP1. Percentage profit is? *74* (1)

In this version of the question steps 1.1 and 1.2 carry 0.5 marks each and the key part itself 1 further mark so that again two marks are possible. This time, however, the student has to calculate and submit three answers in order to gain full marks. The benefit in formative assessment is that the student can progress when stuck on how to make the first step. Moreover, with ticks and crosses visible on screen the student can gain confidence as the answers are entered to steps 1.1 and 1.2. In summative testing the device of optional steps provides for some partial credit for even if the student cannot make the final step of calculating that the percentage profit is to be achieved by forming the fraction (actual profit)/(original sale price) = 111/150 expressed as a percentage. In other words express 111/150 as x/100 where x is the number needed as the answer for key part 1. One other variant of this approach would be to provide the student who is stuck with a hint called an information step, which does not seek an answer, and would on the pressing of the optional steps button appear as:

Question 6.10: A greengrocer buys a bag of nine oranges for £1.50 and sells them on at 29 pence each. What is the percentage profit on the sale when all nine oranges have been sold?

Step 1.1. First, calculate how much in pence there is when all nine oranges are sold. (0)
Step 1.2. What is the profit in pence when all nine oranges have been sold? *111* (1)

KP1. Percentage profit is? *74* (1)

This time no marks are given for the information step 1.1 and one mark for step 1.2 with a further mark for the key part 1 itself giving again two marks for completing this question. This is again good practice for formative assessment where the emphasis is to provide feedback and encourage the student to make progress.

It could be argued though that this does not properly reward the good student in a summative exercise. Even though the average student has taken longer to complete this question through the steps and may have put some additional pressure on himself/herself to complete the whole timed test nevertheless the more traditional examiner might wish to differentiate between candidates more subtly. So, finally, consider the example in the form:

Question 6.11: A greengrocer buys a bag of nine oranges for £ 1.50 and sells them on at 29 pence each. What is the percentage profit on the sale when all nine oranges have been sold?

Step 1.1. First, calculate how much in pence there is when all nine oranges are sold. (-0.5)
Step 1.2. What is the profit in pence when all nine oranges have been sold?*111* (0.5)

KP1. Percentage profit is? *74* (2)

This time a more traditional examiner might argue that information step 1.1 gives away some of the strategy for this question so that 0.5 marks should be deducted for this hint. So, when step 1.2 is given 0.5 marks and key part 1 an additional mark then the student can score 2 - 0.5 = 1.5 marks for a correct solution. This approach helps to differentiate between candidates and adds to the sophistication of automatic testing which is a feature currently under investigation by Korabinski et al (in progress).

Summary
There are 4 different ways to present a question. These are:

- Single Key Part
- Multiple Key Parts
- Key Part/s and optional steps
- Key Part/s and optional steps with marks deducted

Section 7: Further Issues of Partial Credit

7.1 Creating partial credit questions in Mathematics

In paper based examinations in mathematics partial credit is normally given to an answer that is not completely correct but nevertheless contains some of the correct ideas. For more details of this the reader is referred to Fiddes et al (2001), McGuire and Youngson (2002) and McGuire et al (2002). As an example of how this can be applied, consider the following question worth 3 marks:

> **Question 7.1:** Find the area of a rectangle with height 8 centimetres and length 7 centimetres.

The correct answer of 56 square centimetres would be awarded 3 marks in either a paper-based or ICT examination. The marking scheme for the paper-based examination may indicate 1 mark for knowing the formula they have to use, 1 mark for substituting the numbers they are given into the formula and 1 mark for doing the arithmetic correctly. Therefore if the following "answer" is given,

Area = height times length
= 8 times 7
= 54

then a candidate in a paper-based examination may well be awarded 2 out of 3 marks for this solution. In an ICT examination, the candidate would get no marks for the answer 54 if only the final answer is marked. This can cause significant evidence of a difference in marks between the same examination taken in different media (see McGuire et al (2002)), the longer and harder the questions the greater the difference. In the 2002 paper, McGuire et al investigated alternative ways of giving an ICT examination to try to overcome the lack of partial credit. Both the methods tried involved the concept of using Steps, that is, breaking the question down into smaller parts that the candidates had to answer. The two methods tried involved compulsory steps and optional steps.

In the first version the software was designed so that candidates had to go through all the steps while in the second the candidates had the choice of taking the steps or not as already described in section 6 above. The marks obtained in either of these tests were comparable to those obtained with partial credit in a paper-based examination. However, even if the marks obtained are similar, this does not mean that the candidates had shown the same skills. In particular, the use of Steps provided the candidate with the strategy to do a question. This is normally a skill that a paper-based examination seeks to test. To determine whether there are other possible differences in skills shown in the two types of examination it is necessary to take a much closer look at a question and marking scheme for its answer. Marks can then be amended for the learning outcomes that are being tested in the question. In the example above about the area of a rectangle, these are:

a) knowing the formula;
b) applying the formula; and
c) doing the arithmetic.

These correspond to each of the lines of the answer shown above and the human marker awards one mark for each one of these learning outcomes performed correctly by the candidate.

This leads to the following approach for translation of a question into an ICT format. Decide what learning outcomes are to be tested (usually indicated by the marking scheme). If one of these is to "know how to" or "know the formula for" set this up as a Step in the question. A candidate has the option to obtain this Step by clicking on a button labelled **<STEPS>.** If they do this then they see some information about how to start the question (this type of step could be called an Information Only Step) together with possibly some additional shorter questions that could guide them to the answer to the overall question. In the example above about the area of a rectangle the Information Only Step may be "The formula for the area of a rectangle is the height times the length". At this point this looks as though it is open to the same type of criticism about the experiment with steps mentioned above but the difference here is that to compensate for the lack of knowledge of one of the learning outcomes full marks should not be awarded when the candidates use Steps as described in section 6.5 above.

In more complicated questions as well as an Information Only Step there could also be other Steps in which candidates have to provide answers. These are awarded marks based on the number of learning outcomes that are being tested. This is very similar to the way that partial credit is awarded by human markers. Therefore using Information Only and other Steps could provide a suitable way of testing the same learning outcomes in a question as those that are currently done in a paper-based examination. In this way, then, this design allows for the possibility of awarding partial credit in an ICT examination.

7.2 Some more Examples

The following two examples illustrate what happened in practice when experimenting with a group of Higher Mathematics students in two Edinburgh schools. Rough working from the ICT examination and candidates' attempts in the paper-based examination were used to find where errors occurred in the candidates' solutions. The statistical analysis showing how the comparison of the performance was done is given in the paper by Korabinski et al (in progress]. The two examples below illustrate how care has to be exercised in choosing the steps. Knowledge on where the students are likely to make their major mistakes is needed when authoring good questions.

> **Question 7.2:** Find the gradient of the tangent to the curve with equation $y = x^2 - 8x + 14$ at the point (6,2). [The question also contained a sketch of the curve, the point (6,2) and the tangent drawn at (6,2).]

The marking scheme gave the learning outcomes as

Know to differentiate to find the gradient:	1 mark;
Find the value of the gradient:	1 mark; and
Write down the equation of the line:	1 mark.

The steps given in this question were

> Step 1.1. To find the gradient of the tangent you first need to find $\frac{dy}{dx}$ (which was an information only Step); and
>
> If $y = x^2 - 8x + 14$ find $\frac{dy}{dx}$ (and here the students were asked to type in their answer).

Then, the key part asked for the equation of the tangent.

So, the good students scored 3 marks usually completing this question without recourse to the **<STEPS>** button. Those that took the steps and received the strategy lost 1 mark but could recover by completing the second step for one mark and gather the final mark by writing down the equation of the line, thus, securing partial credit of 2 out of the possible 3 marks.

In general most errors occurred at the start of the question-usually not realising that differentiation was required to tackle this problem. The actual differentiation and substitution caused little problem and so this question produced no significant difference in marks between ICT with Steps and a paper version of it.

Consider the following option:

> **Question 7.3:** Find the derivative of $y = (x^3-1)/x^2$.

In this question the marking scheme gave the learning outcomes as:

> Know how to divide out a quotient --- 1 mark;
> Differentiate the x term --- 1 mark; and
> Differentiate - $1/x^2$ --- 1 mark.

The steps given in the ICT version of the question were

> To find the derivative you first need to divide x^3 -1 by x^2 (which was the Information Only Step).
> If $(x^3-1)/x^2 = x^p - x^q$, what is the value of p?
> If $(x^3-1)/x^2 = x^p - x^q$, what is the value of q?

(The candidates had to type in the values of p and q.) The key part then asked for the derivative of the whole expression. Three marks were awarded for a correct key part answer and those taking the steps could only gain a little partial credit. For, with one mark lost with the information only step this could be re-gained by giving 0.5 marks for each of p and q. The final answer would have gained one further mark but since the error occurred for most students at the end of the question a zero mark was the most frequent result for those not taking steps and only one mark for those who did. Whereas, on paper many students could gather two out of the three marks on offer.

Most candidates knew that they had to split this into two terms and could differentiate the term x so most of the errors occurred at the end of the question differentiating $1/x^2$. In this case the Steps were of little use. Many of the students took the question without using steps and gave answers in the ICT examination which were wrong (so were awarded no marks) making their errors at the end with an incorrect derivative for $1/x^2$. Those doing a paper-based examination got a lot of partial credit for incorrect answers since their early working was rewarded. This produced a significant difference in marks between ICT with Steps and paper-based examinations.

One point to note here is that if the main learning outcomes are to divide out a quotient and then differentiate a negative power of x it might have been a better policy to test these two outcomes separately in two questions.

7.3 One possible approach.

The result of Question 7.2 shows that in this case Steps were a good way of replicating the partial credit a human marker would give, but this was not the case in Question 7.3. This may look disappointing from the viewpoint of replication of results by ICT examinations from paper-based examinations but there is a way to align the ICT questions so that they better fit the learning outcomes for the original question. What happened was that a pre-existing question from a paper-based examination was changed into an ICT examination question that tried to re-produce the examination of the same learning outcomes as the paper-based question using a similar structure. However, if instead the learning outcomes to be tested were the starting point, then it may be possible to construct one or more questions that test these learning outcomes equally well in both paper-based examinations and ICT examinations. In question 7.3, the learning outcomes were dividing a quotient, differentiating a positive power, differentiating a negative power. These could be examined in two questions as follows.

Question 7.4: Find the derivative of $(x^4-x^3)/x^2$.
Question 7.5: Find the derivative of $1/x^2$.

Either or both of these new questions could have steps. Question 7.4 could have the same steps as the original question while question 7.5 could have the information only step telling the candidates to express $1/x^2$as x^p for example. It is likely that there will still be some discrepancies between marks in certain questions and only by several iterations will a satisfactory outcome be achieved. However it is hoped that the number of iterations can be reduced with the help of the analysis outlined above.

7.4 Follow through errors.

Follow through errors occur in both paper based examinations and ICT examinations although allowances can be made using partial credit in paper based examinations. Consider the following example.

Question 7.6: Find the coordinates of the stationary points of the curve with equation $y = x^3/3 + x^2- 3x - 11$. Using differentiation determine their nature.

The candidate is set two tasks in this question. If they get the answer to the first wrong then they are likely to get the second wrong also even if they have the correct method for determining the nature of the stationary point and make no mistakes in this process. In a paper-based examination the marker would award partial credit for the method used to classify stationary points even if the values the candidates had for the stationary points are wrong. One way round this is to split such a question into two as follows.

Question 7.7: Find the coordinates of the stationary points of the curve with equation $y = x^3/3 - 2x^2+ 3x + 1$.

Question 7.8: The curve with equation $y = x^3/3 - x^2- 3x + 1$ has derivative $x^2- 2x - 3$ and stationary points at (-1, 8/3) and (3, -8). Using differentiation, determine the nature of the stationary points.

The same learning outcomes are examined by the two new questions as the original. The advantage over the original is that even if the candidate does make errors in finding the stationary points these do not affect the candidates' ability to obtain credit for knowing how to classify the stationary points.

7.5 Questions involving proofs.

For some questions the answer is contained within the question, for example,

> **Question 7.9:** Show that the line with equation y = 2x - 3 is a tangent to the circle with equation $x^2 + y^2 + 2x - 4 = 0$

Even in a paper-based examination this can be awkward to mark. A candidate may well write down working unrelated to what is required and then finish off their solution with "therefore the line with equation y = 2x-3 is a tangent to the circle with equation $x^2+ y^2 + 2x - 4 = 0$". Even though that is what is required as the last line for the correct solution a marker has to look carefully at the whole solution to see how much credit should be awarded. This makes this type of question very awkward for ICT assessment. Here are two possible approaches to tackling this.

> **Question 7.10:** Describe how to show that the line with equation y = 2x - 3 is a tangent to the circle with equation $x^2+ y^2 + 2x - 4 = 0$.
>
> Find the x-coordinate of the point of intersection of this line and circle.

The student is first asked to complete the sentence "To show that this line is a tangent to this circle I would" describing the process that they would go through to carry out the question and then asked for the x-coordinate of contact.

In general it is harder to explain how to do something rather than just doing it so that this revised question may be considerably more challenging than the original. However, the description then has to be assessed by a human marker unless some free text marking package is available.

An alternative approach is to turn this into an ICT question using random numbers as follows:

In this case the equation of the line is randomised so that in some cases it is a tangent and in other cases it cuts the circle in two distinct points. The candidates are then asked for the intersection points of the line and the circle and then asked if the line is a tangent or not.

It is not a tangent for the numbers given in question 7.11.

> **Question 7.11:** Determine whether or not the straight line with equation y = 2x + 2 is a tangent to the circle $x^2+ y^2 + 2x - 4 = 0$.

It could be argued that the second version of the ICT question is harder as candidates do not have the reassurance in advance of what the answer is going to be. Nevertheless, in terms of learning outcomes, this may be the closer version of the two ICT questions to the original learning objectives and difficulty of the question.

Either of these versions may work for other similar type questions but there still remain problems with questions that require proofs, for example, or questions that test the understanding of mathematical induction. Generally, these are not testing the lower order skills but rather analysis, synthesis or evaluation. The tools of CAA are generally unable to measure the higher order skills consistently at the present stage of development in subjects like Mathematics.

Section 8: Some Chemistry Examples

8.1 A Higher Chemistry example

Higher Chemistry National Assessment Bank 1 (NAB) called Energy Matters; D069 12/NAB004 was one of the chosen tests in phase 1 of the PASS-IT Project. Consider a question from that test:

Question 8.1: Study the following data, which relate to the chlorides of the second period elements.

Chloride	LiCl	$BeCl_2$	BCl_3	CCl_4	NCl_3	OCl_2	FCl
Mp/°C	614	405	-107	-23	-27	-20	-154
Bp/°C	1350	487	12	77	71	4	-101

(a) (i) Which chloride in the table is a gas at 0°C? (1)

The answer to the above question part is one of the 7 Chlorides given in the table, specifically FCl.

This was authored as free text entry, marked with word match, allowing the student to enter any textual answer they desired. This then requires very careful consideration of the permitted answers. Do you allow FCL, with incorrect capitalisation? Even if this capitalisation would not be accepted on paper, could this be considered a typing error due to using an unfamiliar medium? Is correct capitalisation an objective that is being measured here (even though the correct capitalisations are shown above)? Do you accept spaces between the letters, such as F Cl, or F C l? Do you mark as correct answers using the element names rather than the symbols, as in this case Fluorine Chloride? Are spelling and capitalisation important here?

The discussions and decisions are all a critical part of the question specification. Another option in this case could be to author the question in a multiple choice format. This would remove all problems associated with input entry. Of course this may not be acceptable if you want to assess their ability to capitalise the chemical formula correctly!

Another issue arose due to the wording of the question. Given that the question asks “which chloride” it would seem that Fluorine by itself would also be a possible answer --- the Chloride is Fluorine. Most of this confusion could have been avoided with a slightly different wording of the question (by employing the word “compound” rather than “chloride”). The other alternative correct answers could have been added by the question setter at the authoring stage.

However, despite all the potential problems with a free text answer of this sort and the associated word match marking scheme, it is interesting to note that the alternatives noted above were only present in a small number of submitted answers in the Chemistry pilots from the candidates in the six Scottish schools taking part in the trials.

8.2 A second example

Another question asked students to “Calculate the volume, in litres, of nitrogen dioxide produced” In a chemical reaction. This question required the students to work out a number of litres and the prompt stated clearly that the units were not needed. However, the question had some students adding the unit despite the careful wording of the prompt. This led to 14 students out of 50 being marked wrong when they were correct in one part of the question and 8 students out of 50 being marked incorrectly on the second part of the same question for the same reason. Specifying the answer in litres was not something that had appeared in the original paper based question, but had been added to reduce the extent of correct answers (so that ml, cm^3 etc, were specifically not considered to be correct).

The desired answer was then simply the number of litres, and had been authored to allow the number to be entered in a variety of mathematically equivalent formats, such as 8, 8.0, 4*2 etc. However, this meant that if a student included the units (litres) in the answer, they would be marked wrong. This kind of confusion is not surprising since students are well drilled to include units in their answers on paper and they may not have practised enough on computer to realise that this would cause the computer a problem.

This is a difficult decision, are we marking the numerical part of the answer, or the presence of correct units, or both? If knowledge of both correct number and correct units is desired, then the student could have been asked for the numeric part and the unit separately, but would that be giving away information that a unit was required? Gap fill questions provide one possible solution to this difficulty. However, some experts would still argue that such a gap would be giving away too much information to the candidate.

Instances of some of the issues that arose and are discussed here could be further reduced, or eliminated, by alternative question design. But, this does highlight the importance of post moderation of student answers and feeding this back into the question design process.

Section 9: A Computing Example

The Higher National Unit Computer Architecture (D75P 34) was one of the units chosen from Computing to pilot in phase 1 of the Pass-IT project. The following example is a question from Outcome 1 of this unit. It is part of the Mandatory Activity for the conversion between binary and decimal representations.

> **Question 9.1:** You are involved in the testing department of a project that is building digital thermometers. The hardware reports temperatures using a 12 bit fixed point binary number, with the most significant bit being used as a sign bit and the lowest five bits being used to hold the fractional part of the reading.
>
> As part of the testing strategy you are required to convert the following decimal number into its corresponding binary representation so that the observed result can be compared against the expected result.
>
> Key Part Prompt: What is the fixed-point binary representation of the decimal 28.65625?
>
> Answer: 0011100.10101

This was authored using a numerical key part and involves randomised parameters to create a more valuable question. The positive value (in both decimal and binary form) is made up from 11 randomised parameters, a to k, each equal to either 0 or 1. A slight restriction on these random parameters is needed to ensure that the value is never equal to 0. This was done by ensuring that both the integer and the fractional parts are never equal to zero. This gives a minimum value of 1.03125 and a maximum of 63.96875. There are a few issues with this question that need consideration.

Firstly, using the numerical key part called a JME (Judged Mathematical Expression) makes it possible for mathematically equivalent answers to be assessed and marked by the system. This is an asset of the system but in this case it may cause incorrect answers to be awarded marks. For example, in this question, missing out the leading zeros to give an answer of 11100.10101 instead of 0011100.10101 would be awarded marks as it is mathematically equivalent but it is not an acceptable answer in Computing. To avoid such an answer being marked correctly, restrictions on the length of the student's answer have been applied. The restrictions ensure that the answer is 13 characters long (inclusive of the decimal place) before any marks are awarded.

Another issue arises as the answer in numerical key parts is automatically simplified by the system stripping any unnecessary leading or trailing zeros. In general, this is the required simplification but in this case it might cause problems with the display of the binary value. If the answer is revealed by the student (this feature is only available in 'Help Mode') then they might see the simplified answer of 11100.10101 mentioned above without leading or trailing zeros. This may prove misleading to the student but has now been resolved.

An extension of this question was also authored to include negative values. Randomising this question needed further conditions as the conversion of negative values involves the added consideration of incorporating the method 'twos complement' into the randomised values.

Extensive extra conditions had to be applied to tailor the required effect so the question is slightly more complicated than the question involving positive values only detailed above.

A question for the conversion from binary to decimal was also authored using the same method and random parameters. Both versions provide versatile questions creating varying values and cover the Mandatory Activity from the Computer Architecture Unit.

Section 10: Further Considerations

10.1 Multi-media elements

This section draws on the article by Ashton et al (2003). Consider a question from the National Assessment Bank for Higher Mathematics as set by the Scottish Qualifications Authority. The student is asked to show the effect of a specific operation on a function. The traditional way of exhibiting appropriate understanding is to draw the transformed function given a diagram of the initial function.

Diagrams 1 and 2 on the worksheet show part of the graph y=f(x).

(a) On Diagram 1, draw the graph of y=-f(x).

(b) On Diagram 2, draw the graph of y=f(x+3).

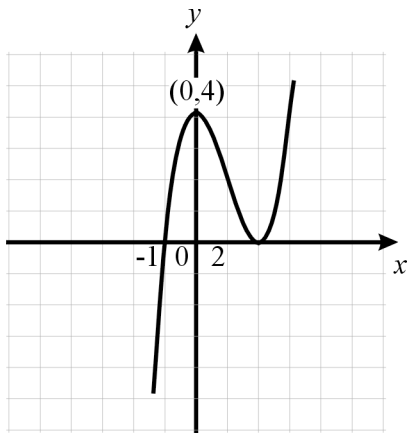


Diagram 1

Figure 1: Original paper based question.

A traditional examiner would be looking for the student to have sketched the original graph reflected in the x-axis. One option for an on-line version would be to use multiple choice, presenting the student with a number of options, possibly combinations of the reflection in each axis. However, this reduces the difficulty of the question from one of the application of knowledge, to recognition and elimination, deemed wholly unsatisfactory and not equivalent to the traditional way of measuring this understanding.

In this case the solution developed for the PASS-IT phase 1 pilots was to use a hotspot question, where each integer coordinate on the graph was a hotspot. The task then became one of clicking on all the turning points and the points of intersection of the x-axis. A selected hotspot was shown as a red circle, and marks were only awarded for all points being correct. Figure 2 shows the students' view after a correct answer has been given.

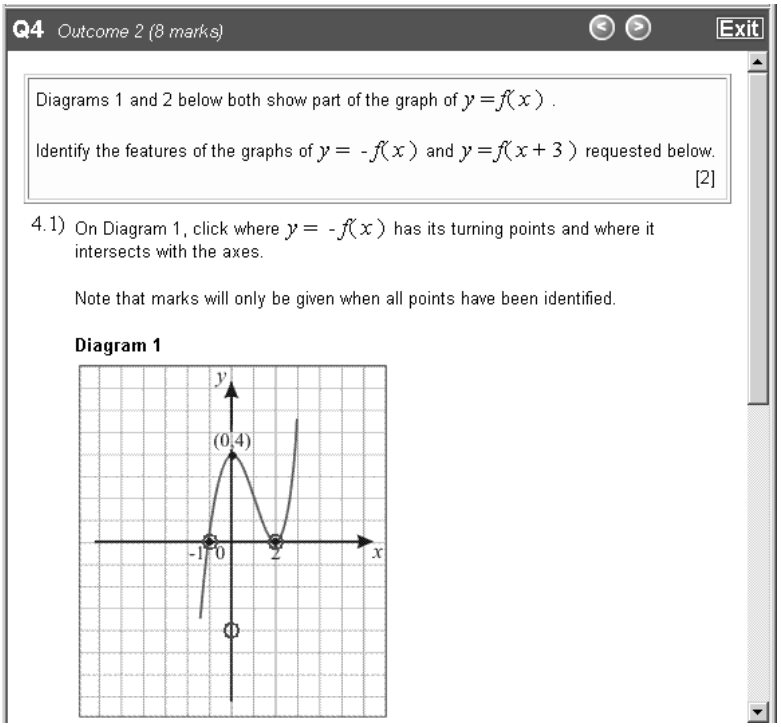


Figure 2: A correct answer to a Hotspot version of the question

This solution removes the problem of only testing recognition as it allows any combination of points to be chosen.

However, ingenious though this solution is, it still poses some problems:

- The student never sees/draws the new function.
- Clicking on the exact point can be difficult, and may reduce accessibility.
- The question takes a long time to produce (this one has 169 hotspots!).
- The question has a large file size, which could pose problems with server load.
- Changes could be tedious, and there is no opportunity for randomisation.

From the students' perspective, not seeing the final function could be a huge barrier to identification of the correct answer, and to learning from that. This problem is likely to be more pronounced in lower school years. In addition, the hotspots themselves must be small so that unique identification can be made, and to keep the image size to an acceptable level. Students with difficulties using the mouse, or where a mouse is "sticky" (a common problem), could have trouble selecting the correct hotspot. Every choice the student makes is communicated to the server, allowing the corresponding highlighted circle to be shown. This takes time, even with the speediest of connections. Combine this with time to deselect incorrectly chosen options, and this question could take a considerable amount of time to answer.

From a technical perspective, a question with 169 hotspots has to contain a reasonable amount of information, increasing the file size quite substantially. As a comparison, this file was ~60KB (not including the image). The next largest file in this NAB unit was only 15KB. Although 60KB is not huge, consider this for even a single class of students, with multiple hotspots to select, and quickly, potential traffic and server load problems may arise.

From an authors' perspective the question takes a significant amount of time to produce, and can be tedious to change. There is also no potential to randomise the function being presented

(something that is useful in formative assessment, and may be useful to combat the close proximity of computers in classrooms during summative testing).

One of the recent technical developments in the assessment engine used in the PASS-IT project has been to implement two-way communication with multimedia elements such as Macromedia Flash™. This enables a new approach to this question (see Figure 3).

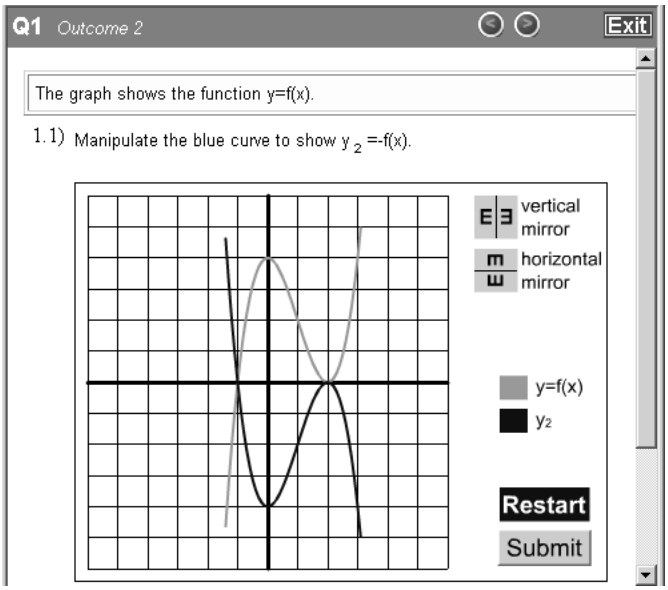


Figure 3: A multimedia version of the question

This solution addresses all of the issues raised above and allows the student to perform translation and reflection operations on the original curve. As with the hotspot solution, the student can position the new function in many locations. The original function is always visible (the lighter of the two curves), however, the student can now also see the new function, improving familiarity and the potential for evaluating their answer and learning from the assessment.

Translations can be performed by using the mouse to drag the curve. The “live” area to drag extends well beyond the original curve, and when “dropped” the curve will snap to the closest grid point, reducing mouse related problems. Reflection operations can be performed using two buttons (top right). In addition, keys can be used for translation (arrow keys) and mirror operations (h and v), further improving accessibility.

From the technical perspective, the combined file size of the Flash™ element and the question is now ~10KB. The answer is only communicated to the server when the student chooses to submit their answer, further reducing traffic and server load.

From the authors’ perspective, once the Flash™ element has been produced, the authoring and modification of questions is dramatically improved. Communication between the assessment engine and Flash™ allows many parameters to be sent in, allowing both extensive reuse, and randomisation of the question.

There are many other examples where appropriate use of integrated technology can enhance the range of potential objectives that can be measured, the flexibility of generation and reuse, and the usability and accessibility of the system. It is hoped that phase 2 of the project will allow further exploration of the use of multimedia.

Flash™ is a trademark of Macromedia, Inc.

10.2 Use of Must have and Not allowed strings

The following example is part of a question that appears in Mathematics Intermediate 2 NAB 1. This is a unit that will be piloted in phase 2 of the Pass-IT project. The question deals with the factorisation of different expressions.

Question 10.1: Factorise the following expressions.

Key Part 1 Prompt: Factorise y^2-5y.

Key Part 1 Answer: *y(y-5)*

Key Part 2 Prompt: Factorise a^2-b^2.

Key Part 2 Answer: *(a+b)(a-b)*

Key Part 3 Prompt: Factorise x^2-x-20.

Key Part 3 Answer: *(x-5)(x+4)*

As this question involves a rearrangement of expressions, each answer is mathematically equivalent to the expression given in the question prompt itself. An assessment system, which can check algebraic expressions, will automatically mark the original expression as correct answers unless the author takes some preventative action. To avoid these expressions simply being reproduced by the student to gain marks, restrictions to acceptable answers must be made.

Using the system, string restrictions to avoid answers containing certain strings being accepted, can be applied. For example, in key part 2 the expression given in the prompt is a^2-b^2 and the answer required is (a+b)(a-b) or (a-b)(a+b). To make sure that one of these answers is the answer given, we must restrict what is allowed in the student's answer. 'Not Allowed' string restrictions are applied so that the strings ^, aa and a*a are not permitted in the student's answer. Zero percent partial credit is assigned to these restrictions so that if any answer contains these strings then 0 marks are awarded.

In addition, what is allowed in the student's answer can be rejected by using 'Must Have' string restrictions. The answer can be constrained so that it must contain a certain string. In the above example, the string restriction 'Must Have' (and) was used. Again, 0% partial credit was assigned to this restriction. This combination of 'Not Allowed' ^, aa and a*a and 'Must Have' (and) ensures that only the correct answer (a+b)(a-b) or (a-b)(a+b) will be awarded any marks. It is possible to mark answers like this by using string-matching techniques but such methods are often cumbersome and lengthy to apply and a student will always come along with a perfectly acceptable mathematical answer that the author had failed to consider. Then, the author has to return to the question and amend it.

Assigning string restrictions to key parts can be useful in many other situations. For example, to ensure a numerical value is rounded to the nearest integer 'Not Allowed'. (a decimal point) can be added. In this case, it may be desirable to award 50% partial credit. This would allow marks to be awarded for the calculation but, without the correct rounding, full marks should not be given. Also, to aid the student, string restrictions can have related messages so that if the restriction is triggered then a message box appears containing, for example, 'Give your answer to the nearest integer'. A similar example occurs when the answer is required as a fraction, say ½. If the

student gives the answer 0.5 then the message “Your answer is correct but not in the correct form” can be supplied and a partial credit mark awarded.

10.3 Looking ahead

It has been predicted that in the future "simple multimedia exercises will give way to virtual reality simulations. Such simulations will model environments-science labs, field experiences - giving students a chance to learn and be assessed under conditions similar to those encountered by practitioners" (Bennett, 1998). Indeed, Milligan and Thomas (2003) recently reported on the early stages of simulation and assessment integration within the context of Chemistry practicals, work that is set to continue in phase 2 of the PASS-IT Project.

Bibliography

Ashton, H.S., Schofield, D.K. and Woodger, S.C. (2003) "Piloting Summative Web Assessment in Secondary Education", *7th International CAA Conference,* Loughborough University, pp.19-29.

Baldwin, B.A., and Howard, T.P. (1983) "Intertopical sequencing of examination questions: an empirical evaluation." *Journal of Accounting Education,* pp.89-96.

Beevers, C., Cherry, B., Foster M. and McGuire, G. (1991) "Software Tools for Computer Aided Learning in Mathematics", Ashgate Publishing Company.

Beevers, C., Foster, M., McGuire G and Renshaw, J. (1992) "Some Problems of Mathematical CAL", *J Comp Educ,* v18, pp119 – 125.

Beevers, C., McGuire, G., Stirling G. and Wild, D. (1995) "Mathematical Ability Assessed by Computer", *J Comp Educ,* 25, pp123 - 132.

Beevers, C, Youngson, M, McGuire, G, Wild D and Fiddes, D, (1999) "Issues of Partial Credit in Mathematical Assessment by Computer", *ALT-J,* v7, pp26-32.

Bennett, R. E. (1998) "Reinventing Assessment: Speculations on the Future of Large-scale Educational Testing" Princeton: Educational Testing Service.

Brown, G., with Bull, J., and Pendlebury, M. (1997) "*Assessing Student Learning in Higher Education",* London: Routledge.

Fiddes, D. J., Korabinski, A.A., McGuire, G.R., Youngson, M.A., and MacMillan, D. (2001) "Are Examination Results affected by the mode of delivery?" *Alt-J*, v10, pp.60 - 69.

Gronlund, N E, (1988) "*How to construct Achievement Tests"* Englewood Cliffs: Prentice Hall

Gruber, R.A. (1987) "Sequencing Exam Questions Relative to Topic presentation" *Journal of Accounting Education,* pp.77-86.

Heard, S., Nicol, J. and Heath, S. (1997) *"Setting effective objective tests"* University of Aberdeen, Geography.

Isaacs, G. (1994) "About multiple choice questions", *Multiple Choice Testing: Green Guide,* No. 16, HERDSA, Campbeltown, NSW, pp.4-22.

Kniveton, B.H. (1996) "A Correlation Analysis of Multiple-choice and Essay Assessment Measures" *Research in Education*, v56, pp.73-84.

Korabinski, A.A., Youngson, M.A. and McAlpine, M. (in preparation) "Can CAA examinations results give more information than traditional examination?"

Matthews, J. (1981) "*The Use of Objective Tests"* Lancaster: University of Lancaster Press. (Revised ed.)

McDonald, A.S. (2002) “The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments” *Computers and Education,* v39, pp.299-312

McBeath, R. (1992) “*Instructing and Evaluating Higher Education: A Guidebook for Planning Learning Outcomes*” ETP: New Jersey.

McGuire G.R., and Youngson, M.A. (2002) “Assessing ICT Assessment in Mathematics” http://ltsn.mathstore.ac.uk/articles/maths-caa-series/mar2002/index.htm

McGuire, G.R., Youngson, M.A., Korabinski A.A. and McMillan, D. (2002) “Partial credit in mathematics exams-a comparison of traditional and CAA exams” *6th International CAA Conference*, Loughborough University, pp.223-230.

Milligan, C. and Thomas R. (2003) Online Assessment of Practical Experiments, *7th International CAA Conference*, Loughborough University, pp.421-430.

Newstead, S. and Dennis, I. (1994) "Examiners examined: the reliability of exam marking in psychology", *The Psychologist*, v7, pp.216-219

Owen, S.V. and Freeman, R.D. (1987) "What's Wrong with Three Option Multiple Items?" *Educational & Psychological Measurement*, v47, pp.513-22.

Paxton, M. (1998) "A Linguistic Perspective on multiple choice questioning" *The Higher Education Research and Development Society of Australasia newsletter*, April 1998 20.1

Stout, D.E. and Wygal, D.E. (1994) “Additional Empirical Evidence on the Relationship Between Exam Sequencing and Accounting Student Performance” *Advances in Accounting* 8, pp.133-152.

Question Mark and GES, (1995) “*Better Testing: A Computerised Guide to Writing Objective Tests*” Question Mark Computing Ltd.

Appendix A: Sample multiple-choice questions

Psychology

1. Episodic binge eating, often with voluntarily induced vomiting, is called

   (a) Aversive appetite
   (b) Anorexia
   (c) Bulimia *
   (d) Appetitive psychosis
   (e) Affective hysteria

2. The therapeutic method of systematic desensitization is most often used in treating

   (a) Schizophrenia
   (b) Autism
   (c) Character disorders
   (d) Depression
   (e) Phobias *

Questions 3 and 4 are based on the following:

In an experiment, subjects were given 1,000 trials per day for five days. On 500 random trials each session, a weak tone of constant intensity was presented after a warning light; on the remaining trials, no tone occurred after the warning. Subjects were required to respond "yes" if they thought a tone had been presented and "no" otherwise. At the beginning of each session, they were instructed as to how much money they could win for correctly reporting the presence or absence of the tones and how much money they would lose for incorrect responses. The amount that could be won or lost varied from session to session.

3. This experiment was probably performed within the framework of

   (a) Signal-detection theory *
   (b) Gestalt theory
   (c) Adaptation-level theory
   (d) The place theory of audition
   (e) The duplicity theory

4. The theoretical and empirical approach exemplified by this experiment has had a major impact on psychophysics because it

   (a) Demonstrated the existence of a sensory threshold
   (b) Suggested that internal noise is absent in sensory systems
   (c) Stressed the effects of adaptation
   (d) Provided a means to separate sensory and motivational variables *
   (e) Validated the power law

Above questions from GRE-Psychology Test descriptive booklet (1997-1999)
http://www.ets.org/practice_test/gre/online.htm

Sociology

1. Studies of the political correlates of social mobility suggest that the upwardly mobile tend to

   (a) Retain their political ties with their class of origin
   (b) Acculturate politically toward the class of destination *
   (c) Become politically radical
   (d) Maintain their previous political orientation
   (e) Decrease their political activity and involvement

2. From his comparison of Chinese, Indian and Western European societies, Max Weber inferred that the differences in their receptivity to capitalism could be accounted for in part by

   (a) Differing systems of religious values *
   (b) Differing political systems
   (c) Differing ecclesiastical structures of churches
   (d) Differential access to raw materials
   (e) The desire of merchants to make a profit

3. A scale that ranks positions on a continuum but does not indicate distance between one position and another in equal units is called

   (a) An interval scale
   (b) A ratio scale
   (c) A nominal scale
   (d) A qualitative scale
   (e) An ordinal scale *

Questions above are from the GRE Sociology Test descriptive booklet (1997-1999)

Economics

1. If the level of nominal net national product is rising during a period in which the money supply is constant, which of the following must also be true?

   (a) Prices are rising
   (b) Prices are constant
   (c) The velocity of money is rising *
   (d) The velocity of money is constant
   (e) Real net national product is rising

Questions 2 and 3 refer to the following macro model of a closed economy. All data are in billions of dollars

$$Y = C+I+G$$
$$C = 0.8 (Y – T)$$
$$I = 40+0.1Y$$

where

Y = income
C = consumption
I = investment
G = government expenditures
T = taxes

2. If there are no taxes or government expenditures, what is the equilibrium level of national income?

   (a) $60 billion
   (b) $100 billion
   (c) $400 billion *
   (d) $500 billion
   (e) $800 billion

3. Assume that the government levies an income tax of 25 percent and simultaneously undertakes expenditure in an amount that balances the budget. What is the equilibrium level of national income?

   (a) $200 billion
   (b) $400 billion
   (c) $600 billion
   (d) $800 billion *
   (e) $1,200billion

4. Assume that a proportional tax is levied on the earnings from the work of a person and that leisure is a normal good. Which of the following best describes the impact of this tax on the person's work hours?

   (a) The substitution effect tends to increase work hours
   (b) The income effect tends to increase work hours *
   (c) Both the substitution effect and income effects tend to decrease work hours
   (d) The substitution effect tends to increase work hours, whereas the income effect tends to decrease work hours so that the net effect is unknown
   (e) The income effect generally dominates the substitution effect, so that the net work hours will decrease

5. Which of the following did NOT contribute to the eighteenth-century Industrial Revolution in Britain?

   (a) Rapid technological change in the textile industry
   (b) Enclosure and consolidation of agricultural lands
   (c) Introduction of gold as a convenient medium of exchange *
   (d) Increased trade with other nations
   (e) Invention and application of the steam engine

Above questions are from the GRE Economics descriptive booklet (1997-1999).

Problem-solving and reasoning for potential law students

The following are sample questions from the American LSAT (Law School Admissions Test), which aims to evaluate the ability of law school candidates to reason and use logic in solving problems:

1. Train service suffers when a railroad combines commuter and freight service. By dividing its attention between its freight and commuter customers, a railroad serves neither particularly well. Therefore, if a railroad is going to be a successful business, then it must concentrate exclusively on one of these two markets.

   For the argument to be logically correct, it must make which one of the following assumptions?

   a. Commuter and freight service have little in common with each other.
   b. The first priority of a railroad is to be a successful business.
   c. Unless a railroad serves its customers well, it will not be a successful business. *
   d. If a railroad concentrates on commuter service, it will be a successful business.
   e. Railroad commuters rarely want freight service as well.

Questions 2-6

A small software firm has four offices, numbered 1,2,3, and 4. Each of its offices has exactly one computer and exactly one printer. Each of these eight machines was bought in either 1987, 1988 or 1989. The eight machines were bought in a manner consistent with the following conditions:

- The computer in each office was bought either in an earlier year than or in the same year as the printer in that office.
- The computer in office 2 and the printer in office 1 were bought in the same year.
- The computer in office 3 and the printer in office 4 were bought in the same year.
- The computer in office 2 and the computer in office 3 were bought in different years.
- The computer in office 1 and the printer in office 3 were bought in 1988.

2. If the computer in office 3 was bought in an earlier year than the printer in office 3 was, then which one of the following statements could be true?

   (a) The computer in office 2 was bought in 1987.
   (b) The computer in office 2 was bought in 1988. *
   (c) The computer in office 4 was bought in 1988.
   (d) The printer in office 4 was bought in 1988.
   (e) The printer in office 4 was bought in 1989.

3. Which one of the following statements could be true?

   (a) The printer in office 1 was bought in 1987.
   (b) The computer in office 2 was bought in 1987.
   (c) The computer in office 3 was bought in 1989.
   (d) The printer in office 4 was bought in 1988. *
   (e) The printer in office 4 was bought in 1989.

4. If as few of the eight machines as possible were bought in 1987, then what is the exact number of machines that were bought in 1987?

   (a) 0 *
   (b) 1
   (c) 2
   (d) 3
   (e) 4

5. If the computer in office 4 was bought in 1988, then which one of the following statements must be true?

   (a) The printer in office 1 was bought in 1988.
   (b) The printer in office 1 was bought in 1989. *
   (c) The computer in office 2 was bought in 1988.
   (d) The computer in office 3 was bought in 1987.
   (e) The printer in office 4 was bought in 1989.

6. Which one of the following can be inferred from the passage?

   (a) If a business does not introduce variety into its computer operating systems, it will lose data on its computers because of damage from virus programs.
   (b) If a computer virus program is introduced into a business's computer, all of the data on that computer will be destroyed.
   (c) If a business introduces variety into its linked computer operating systems, it will have increased overall protection for its systems, but will not have protected every computer from viral invasion. *
   (d) If a business does not have multiple, linked computer systems, its computers cannot be protected from computer viruses.
   (e) If minor variations are created in computer operating system software, it will be easier to access the data on the computers that use that software.

From LSAT Sample Test, June 1991

Appendix B - Online Resources

Question Mark Computing - http://www.questionmark.com/us/home.htm

Educational Testing Service network: http://www.ets.org/
Practice tests: http://www.ets.org/prep.html

TACO Project –UCL (Latin language scansion exercises)
http://taco.cs.ucl.ac.uk:8080/taco/www/showcase.html

Engineering questions - Loughborough University
http://www.lboro.ac.uk/faculty/eng/engtlsc/Eng_Mech/tutorials/tut_index.htm

Multiple Choice Questions, Examinations, and Test Banks - Xavier University
http://www.xula.edu/~jsevenai/objective/objective.html

Improving Your Test Questions - University of Illinois
http://www.oir.uiuc.edu/dme/exams/ITQ.html

Designing and Managing Multiple Choice Questions - University of Cape Town
http://www.uct.ac.za/projects/cbe/mcqman/mcqman01.html

Mathletics - Martin Greenhow, Brunel University
http://www.brunel.ac.uk/~mastmmg/

Marketing Learning Centre at the University of Ulster
http://www.busmgt.ulst.ac.uk/h_mifflin/

CASTLE project - University of Leicester
http://www.le.ac.uk/cc/ltg/castle/

CVU Assessment Engine
http://cvu.strath.ac.uk/ae/index.html

TRIADS - University of Derby
http://www.derby.ac.uk/assess/talk/quicdemo.html

POETICA - Poetry analysis - University of Sunderland
http://www.sunderland.ac.uk/~us0cma/poetica.html

MEDWeb - University of Birmingham
http://medweb.bham.ac.uk/caa/mcq/

Test and Learn - University of Bristol
http://www.tal.bris.ac.uk/

"Web + QMark + Humanities = ?" by Chris Hopkins
http://info.ox.ac.uk/ctitext/publish/comtxt/ct16-17/hopkins.html

The following sites contain examples of mathematical questions and questions that use optional steps:
http://www.calm.hw.ac.uk/

http://www.i-assess.co.uk