

Using multiple choice questions effectively in Information Technology education

Karyn Woodford and Peter Bancroft

*School of Software Engineering and Data Communications
Queensland University of Technology*

As academics are confronted with problems such as larger classes and the introduction of a trimester year of study, it has become increasingly necessary to search for alternative forms of assessment. This is certainly the case in Information Technology (IT), where more lecturers are using multiple choice questions as a matter of expediency and in some instances, the quality of the assessment is being neglected. This paper provides guidance for IT lecturers who wish to write effective tests containing good multiple choice questions. Some of the points raised are founded in the long history of research into this form of assessment but IT lecturers are, in general, unlikely to be familiar with many of the matters discussed. The paper also considers the major criticism of multiple choice questions (that they do not test anything more than just straight recall of facts) and examines ways of overcoming this misconception. It is our aim to raise awareness of these issues in IT education, but teachers in other disciplines may also find the material useful.

Keywords: large classes, IT, multiple choice tests, multiple choice questions

Introduction

Each year growing numbers of students are enrolling in tertiary studies and teachers are facing increased time pressures in setting and marking assessment items. In Information Technology, even with recent downturns in enrolments, teachers of first year units are still faced with classes in excess of 250 students. In other disciplines, alternative testing techniques such as multiple choice questions have long been used to help alleviate the problem and there is an awareness of the extensive body of research in the area. With IT lecturers now making widespread use of multiple choice questions (Lister, 2001; Carter, Ala-Mutka, Fuller, Dick, English, Fone & Sheard, 2003), this paper will provide some guidelines to assist in the construction of well written questions. In addition to the belief that multiple choice questions are often not well constructed (Paxton, 2000), this form of assessment still faces criticism due to the belief that it does not test anything deeper than a superficial memorising of facts. We contend that it is possible to construct multiple choice questions that are able to test higher levels of cognition. The following diagram represents the levels within the cognitive domain as identified by Bloom (1956). The simple recall of facts is at the lowest level, increasing to the evaluation skills at the top.

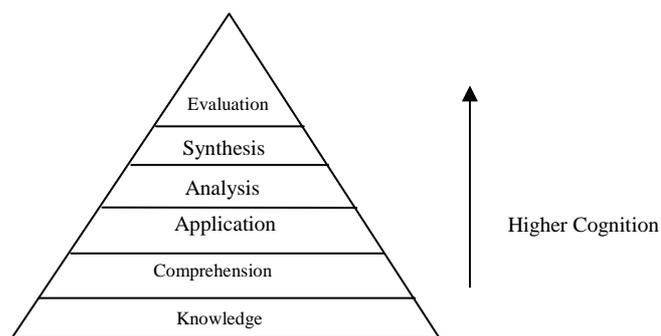


Figure 1: Bloom's levels of cognition

In the paper we address the problem of how multiple choice questions can test more than just knowledge of a subject. Specifically we discuss the comprehension, application and analysis levels of cognition, and give examples of multiple choice questions to test students at these levels.

Firstly we review the terminology used to describe multiple choice questions and suggest methods for measuring their effectiveness, discussing a range of factors that should be considered when composing questions, such as:

- Is the grammar and wording of a question correct?
- How many alternative answers should be used?
- Should negatives and double negatives be used in questions?
- Is it valid to have multiple correct answers?
- How should the questions be ordered (by topic or randomly)?
- Should the alternative answers be listed in any particular order?
- Should 'All of the above' or 'None of the above' be used?
- Are the distracters plausible?

The contribution of this paper is that it will provide guidance for IT teachers who want to set multiple choice questions while maintaining the integrity of their assessment.

Writing effective multiple choice questions

The parts of a multiple choice question

Common terminology (Isaacs 1994) for describing the separate parts of a multiple choice question is illustrated in the following example:

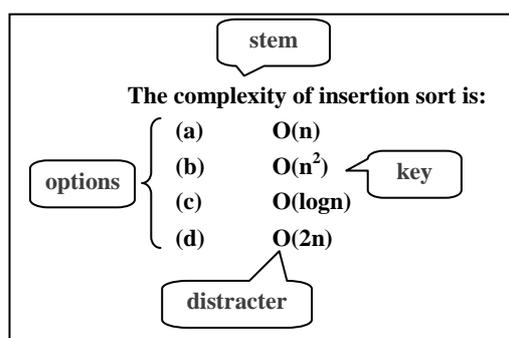


Figure 2: The parts of a multiple choice question

A single multiple choice question, such as the one above, is known as an *item*. The *stem* is the text that states the question, in this case 'The complexity of insertion sort is'. The possible answers (correct answer plus incorrect answers) are called *options*. The correct answer (in this case b) is called the *key*, whilst the incorrect answers (a, c and d) are called *distracters*.

What is an effective question?

A simple measure of the effectiveness of a question is provided by the distribution of student responses amongst the options. If too many students select the correct answer, then perhaps the distracters are not convincing. If very few students answer correctly, then the question may not be clear or a deliberately misleading distracter may have been used. The proportion of students answering a question correctly is called its *facility*. Whilst there are no hard and fast rules about an item's facility, it may be appropriate to have a range somewhere between 0.4 and 0.6 when the goal of the test is to rank students in order. A facility of 0.8 or higher may be more appropriate for tests which are primarily formative (Isaacs, 1994). Another measure of a question's effectiveness is whether the question tests the desired level of cognition (as described in Figure 1 above).

Limitations of multiple choice questions

The traditional style of multiple choice questions - a simple stem question with a key and distracters - has its limitations. A student may select the correct answer by knowing that answer is correct or by eliminating all of the other options. While this may initially seem acceptable, it does not necessarily test the students' full knowledge of the subject - knowing one option is correct does not guarantee they know

that the others are incorrect. Similarly, working out the correct answer by a process of elimination does not demonstrate that the student necessarily knows the solution - if faced with that single answer in a true / false environment, they may not have known that it was correct. This limitation may be overcome by using a different style of multiple choice question (see later).

Factors affecting the validity of multiple choice questions

When writing good multiple choice questions there are several factors to consider - some relate to the actual question whilst some relate to the options (key and distracters).

Correct grammar and wording

The use of incorrect grammar in the stem of a question can often allow students to exclude an option immediately. Consider the following question.

The code fragment 'char * p' is a way of declaring a

- (a) pointer to a char.
- (b) array of strings.
- (c) pointer to a char or an array of strings.

A test wise student may identify option (b) as being incorrect as it starts with a vowel and the stem ends with 'a' and not 'an'. To avoid cueing a student in this manner, the options should include the article:

The code fragment 'char * p' is a way of declaring

- (a) a pointer to a char.
- (b) an array of strings.
- (c) a pointer to a char or an array of strings.

There are several other grammatical considerations (Wilson & Coyle, 1991):

- ensuring the stem and options are worded in the same tense;
- avoiding additional qualifying words or phrases to the key (a test wise student will often identify a longer, more precise answer as the correct option); and
- using similar wording in all options, particularly making sure that the key does not sound like it is directly from a text book.

Number of options

The number of options is one of the most fiercely debated issues amongst supporters of the multiple choice question. Strong arguments have been made for 3, 4 and 5 options. Those who argue for 5 option tests believe that 3 or even 4 option tests increase the probability of a student guessing the correct answer to an unacceptably high level. Those who argue for 3 option tests claim that their tests can be as effective as a 4 or 5 option test, as the additional distracters are likely to be less believable. The arguments for 3 option and 4 or 5 option tests are considered below, along with a brief discussion on removing non-functioning items. Once the number of desired options is decided, it is advisable to use this number of options for every item in the examination to reduce the possibility of careless mistakes.

Three options

A well written multiple choice question with three options (one key and two distracters) can be at least as effective as a question with four options. According to Haladyna and Downing (1993) roughly two thirds of all multiple choice questions have just one or two effectively performing distracters. In their study they found that the percentage of questions with three effectively performing distracters ranged from 1.1% to 8.4%, and that in a 200 item test, where the questions had 5 options, there was not one question with four effectively performing distracters.

The argument for three options therefore is that the time taken to write third and possibly fourth distracter (to make a 4 or 5 option test) is not time well spent when those distracters will most likely be ineffective. In Sidick and Barrett (1994) it is suggested that if it takes 5 minutes to construct each distracter, removing

the need for a third and fourth distracter will save ten minutes per question. Over 100 questions, this will save more than 16 hours of work. Supporters of 4 or 5 option tests would argue that any time saved would be negated by a decrease in test reliability and validity. Bruno and Dirkwager (1995) find that, although reliability and validity are improved by increasing the number of alternatives per item, the improvement is only marginal for more than three alternatives.

Four or five option

The most significant argument against three option multiple choice tests is that the chance of guessing the correct answer is 33%, as compared to 25% for 4 option and 20% for 5 option exams. It is argued that if effective distracters can be written, the overall benefit of the lower chance of guessing outweighs the extra time to construct more options. However, if a distracter is non-functioning (if less than 5% of students choose it) then that distracter is probably so implausible that it appeals only to those making random guesses (Haladyna & Downing 1993).

Removing non-functioning options

Removing a non-functioning distracter (i.e. an infrequently selected one) can improve the effectiveness of the test. In Cizek and O'Day (1994) a study of 32 multiple choice questions on two different papers was undertaken. One paper had 5 option items, whilst the other paper contained 4 option items, a non-functioning item from the identical 5 option item having been removed. The study concluded that when a non-functioning option was removed, the result was a slight, non-significant increase in item difficulty, and that the test with 4 option items was just as reliable when compared to the 5 option item test.

'Not' and the use of double negatives

Asking a student to select which option is not consistent with the stem can be an effective test of their understanding of material but teachers should be very careful when using 'not' to ensure that it is very obvious to the student. A student who is reading too quickly may miss the 'not' keyword and therefore the entire meaning of the question. It is suggested that when 'not' is used, it should be made to stand out, with formatting such as bold, italics or capitals.

Whilst the use of 'not' can be very effective, teachers should avoid the use of double negatives in their questions, as it makes the question and options much more difficult to interpret and understand.

Multiple correct answers

As discussed previously, multiple choice questions have some limitations - specifically that a student may be able to deduce a correct answer, without fully understanding the material. Having multiple correct answers helps eliminate this issue but it is generally agreed that multiple choice questions with more than one key are not an effective means of assessment.

A hybrid of the multiple answer and the conservative formats can be achieved, by listing the 'answers' then giving possible combinations of correct answers, as in the following example:

Which of the following statements initialises x to be a pointer?

(i) `int *x = NULL;` (ii) `int x[] = {1,2,3};` (iii) `char *x = 'itb421';`

- (a) (i) only
- (b) (i) and (ii) only
- (c) (i), (ii) and (iii)
- (d) (i) and (iii) only

In this format the student has to know the correct combination of answers. There is still a possibility that if they know one of the answers is incorrect then this may exclude one (or more) options, but by applying this hybrid format, a more thorough test of their knowledge is achieved.

Order of questions

At issue here is whether questions should be in the same order as the material was taught, or scrambled. In Geiger and Simons (1994), the results of studies indicate that the ordering of questions makes no difference to the time taken to complete the examination, or to the results but it may have an effect on

student attitude. The authors suggest that the reason why question ordering does not have much impact is that most students seem to employ their own form of scrambling, answering the questions they are confident with, and going back to others later.

Order of options

It is recommended that options be arranged in some logical pattern – however, patterns among the keys within a multiple choice test should be avoided (for example, having a repeating ABCD sequence). To ensure that there is no pattern to the keys, it might be advantageous to apply some kind of constraint on the options (for example, put them in alphabetical order) (Wilson & Coyle, 1991).

Use of all of the above and none of the above

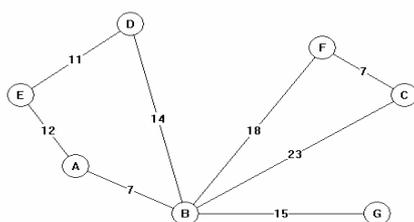
The option ‘all of the above’ should be used very cautiously, if not completely avoided. Students who are able to identify two alternatives as correct without knowing that other options are correct will be able to deduce that ‘all of the above’ is the answer. In a 3 option test this will not unfairly advantage the student but in a 4 or 5 option test a student may be able to deduce that the answer is ‘all of the above’ without knowing that one or even two options are correct. Alternatively, students can eliminate ‘all of the above’ by observing that any one alternative is wrong (Hansen & Dexter, 1997). An additional argument against the use of ‘all of the above’ is that for it to be correct, there must be multiple correct answers which we have already argued against.

The use of ‘none of the above’ is more widely accepted as an effective option. It can make the question more difficult and less discriminating, and unlike ‘all of the above’, there is no way for a student to indirectly deduce the answer. For example, in a 4 option test, knowing that two answers are incorrect will not highlight ‘none of the above’ as the answer, as the student must be able to eliminate all answers to select ‘none of the above’ as the correct option.

In Knowles and Welch (1992) a study found that using ‘none of the above’ as an option does not result in items of lesser quality than those items that refrain from using it as an option.

Writing plausible distracters

An important consideration in writing multiple choice questions is that the distracters are plausible. Poorly written distracters could easily cue a student to the correct answer. For example, if a question asked a student,



Given this undirected graph, what would be the result of a depth first iterative traversal starting at node E?

- (a) EABCFDG
- (b) EDBFCG
- (c) EDBGFCA
- (d) EADBCFG
- (e) EGDCFBA

certain distracters would be ineffective - a distracter that did not include every node would be clearly wrong (option b). Most students would also realise that the second node in a traversal would usually be one close to the starting node, so writing an option that jumps suddenly to the other ‘end’ of the graph may also be easily discarded (option e).

When writing distracters for this question, a teacher should consider the types of mistakes associated with a poor understanding of the algorithm and attempt to offer distracters that include these errors. Additionally, an option containing the answer to a similar type of question could be a good distracter - for example, in this traversal question a distracter could contain the correct result for a depth first recursive traversal (option a) or a breadth first traversal (option d). Only a student who knows the correct algorithm and is able to apply it to the graph will be able to determine which of the plausible options (a, c or d) is the actual key.

Testing more than just recall

The main advantage of multiple choice tests is obvious - they result in a significant reduction of marking for teachers. One of the greatest criticisms of using this type of questioning is that it only tests facts that students can learn by rote. An extension of this argument is the contention that whilst multiple choice questions may be useful for formative assessment and perhaps even mid-semester examinations, they have no place in examinations where the student should be tested on more than just their ability to recall facts. We believe that well written multiple choice questions can test up to the sub-synthesis levels of cognition, that is, knowledge, comprehension, application and analysis. It should be noted that whilst we are arguing in favour of using multiple choice questions to test more than just recall, there is always a place for testing knowledge, including fundamental facts that every student of a subject should know.

Testing comprehension

To test students at the comprehension level, we should present questions that require them to understand information, translate knowledge into a new context, interpret facts and predict consequences. In IT, we could ask students to interpret code or predict the result of a particular change to a data structure, for example, as in the following question:

A **minimum** heap functions almost identically to the maximum heap studied in class – the only difference being that a minimum heap requires that the item in each node is **smaller** than the items in its children. Given this information, what method(s) would need to be amended to change our implementation to a minimum heap?

- (a) insert() and delete()
- (b) siftUp() and siftDown()
- (c) buildHeap
- (d) none of the above

This question tests that the student understands the implementation of the maximum heap, and also asks them to translate some pre-existing knowledge into the new context of a minimum heap.

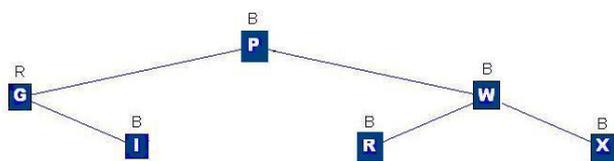
Testing application

The application level requires solving problems by applying acquired knowledge, facts, techniques and rules. To test a student's application of knowledge in a subject, they could be asked, for example, to apply a known algorithm to some data.

In Computer Science subjects there are many opportunities to test at the application level, for example asking the student to apply:

- searching and sorting algorithms,
- ADT-specific algorithms (eg AVL-Tree rotations, Hash Table insertions)
- other algorithms (eg Dijkstra)

The question below tests application of knowledge by asking the student to apply a known algorithm.



Consider the given AVL Tree. What kind of rotation would be needed to rebalance this tree if the value 'H' was inserted?

- (a) no rotation required
- (b) RL
- (c) LR

Testing analysis

Analysis requires the examination of information, breaking it into parts by identifying motives or causes; identifying patterns; making inferences; finding the underlying structure and identifying relationships.

Asking a student to analyse the effect of some code on a given data structure, or identify patterns in the way an ADT processes information are a good way to test their ability to analyse. Asking these questions in a multiple choice format can be very difficult. If you asked a student “What effect does the above code have on our DataSet?” the distracters may give themselves away – the student may easily be able to see that the code is not doing what the distracter claims.

There are a several alternatives to this approach. For example, asking the student whether the code will have the desired effect may allow the writing of more plausible distracters, or alternatively, asking them to analyse some code and then make a comparison with some known code. For example:

Consider the code below which could be used to find the largest element into our sorted, singly linked list called SD2LinkedList. This code would fit one of the processing patterns that we studied in class. Which of the following methods fits the same pattern as this new code?

- (a) union
- (b) hasElement
- (c) exclude
- (d) isSubsetOf

This question not only tests the student’s ability to analyse the new code, but also their knowledge of existing code and their ability to compare the way in which the given code processes data compared to that existing code.

Another method of testing a student’s higher cognitive skills is through the use of linked sequential questions which allows the examiner to build on a concept. An example of this method would be to ask a number of questions each of which makes a small change to a piece of code, and to ask what effect that change would have on the functioning of a program. The student could be required to use the outcome of each question to answer the subsequent question. Using this technique, care needs to be taken to avoid unfairly penalising the student through accumulated or sequential errors.

Conclusion and future work

In this paper, we have attempted to raise the awareness of Information Technology teachers to the vast amount of research that has been undertaken into writing multiple choice questions. We have discussed the terminology used to describe multiple choice questions and their limitations, as well as a range of factors that should be considered when composing questions.

Further, we have described how multiple choice questions can be used to test more than straight recall of facts. We gave specific examples which test students’ comprehension of knowledge and their ability to apply and analyse that knowledge and we suggest that sequentially dependent questions also facilitate testing of higher cognition. Being able to set good questions which test higher cognition allows teachers to use multiple choice questions in end of semester summative tests with confidence, not just as a convenience for low valued mid-semester tests and formative assessment

In other related work, the authors are implementing a web based multiple choice management system. A stand alone prototype of this system (Rhodes, Bower & Bancroft 2004) is currently in use, while the web based system will allow further features, including concurrent access and automatic generation of paper based examinations.

References

- Bloom, B.S. (Ed) (1956). *Taxonomy of educational objectives. Handbook 1. The cognitive domain*. New York: McKay.
- Bruno, J.E. & Dirkwager, A. (1995). Determining the optimal number of alternatives to a multiple-choice test item: An information theoretic. *Educational & Psychological Measurement*, 55(6), 959-966.

- Carter, J., Ala-Mutka, K., Fuller, U., Dick, M., English, J., Fone, W. & Sheard, J. (2003). How shall we assess this? *ACM SIGCSE Bulletin, Working group reports from ITiCSE on Innovation and technology in computer science education*, 35(4), 107-123.
- Cizek, G.J. & O'Day, D.M. (1994). Further investigation of nonfunctioning options in multiple-choice test items. *Educational & Psychological Measurement*, 54(4), 861-872.
- Geiger, M.A. & Simons, K.A. (1994). Intertopical sequencing of multiple-choice questions: Effect on exam performance and testing time. *Journal of Education for Business*, 70(2), 87-90.
- Haladyna, T.M. & Downing, S.M. (1993). How many options is enough for a multiple choice test item? *Educational & Psychological Measurement*, 53(4), 999-1010.
- Hansen, J.D. & Dexter, L. (1997). Quality multiple-choice test questions: item-writing guidelines and an analysis of auditing testbanks. *Journal of Education for Business*, 73(2), 94-97.
- Isaacs, G. (1994). *HERDSA Green Guide No 16. Multiple Choice Testing*. Campbelltown, Australia: HERDSA.
- Knowles, S.L. & Welch, C.A. (1992). A meta-analytic review of item discrimination and difficulty in multiple-choice items using 'None of the Above'. *Educational & Psychological Measurement*, 52(3), 571-577.
- Lister, R. (2001). Objectives and objective assessment in CS1. *ACM SIGCSE Bulletin, ACM Special Interest Group on Computer Science Education*, 33(1), 292-296.
- Paxton, M. (2001). A linguistic perspective on multiple choice questioning. *Assessment & Evaluation in Higher Education*, 25(2), 109-119.
- Rhodes, A., Bower, K. and Bancroft, P. (2004). Managing large class assessment. In R. Lister and A. Young (Eds), *Proceedings of the sixth conference on Australian computing education*. (pp.285-289). Darlinghurst, Australia: Australian Computer Society.
- Sidick, J.T. & Barrett, G.V. (1994). Three-alternative multiple choice tests: An attractive option. *Personnel Psychology*, 47(4), 829-835.
- Wilson, T.L. & Coyle, L. (1991). Improving multiple-choice questioning: Preparing students for standardized tests. *Clearing House*, 64(6), 422-424.

Peter Bancroft, School of Software Engineering and Data Communications, Queensland University of Technology, GPO Box 2434, Brisbane QLD 4001. p.bancroft@qut.edu.au
Karyn Woodford, School of Software Engineering and Data Communications, Queensland University of Technology, GPO Box 2434, Brisbane QLD 4001. k.woodford@qut.edu.au

Please cite as: Woodford, K. & Bancroft, P. (2004). Using multiple choice questions effectively in Information Technology education. In R. Atkinson, C. McBeath, D. Jonas-Dwyer & R. Phillips (Eds), *Beyond the comfort zone: Proceedings of the 21st ASCILITE Conference* (pp. 948-955). Perth, 5-8 December. <http://www.ascilite.org.au/conferences/perth04/procs/woodford.html>

Copyright © 2004 Karyn Woodford & Peter Bancroft

The authors assign to ASCILITE and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to ASCILITE to publish this document on the ASCILITE web site (including any mirror or archival sites that may be developed) and in printed form within the ASCILITE 2004 Conference Proceedings. Any other usage is prohibited without the express permission of the authors.