The Multiple Choice Question in Medicine

John Anderson

Second Edition

The Multiple Choice Question in Medicine

John Anderson
MB, BS, FRCP

Academic Sub-Dean of the Medical School,
and Senior Lecturer in Medicine,
University of Newcastle upon Tyne

Honorary Consultant Physician,
Royal Victoria Infirmary,
Newcastle Area Health Authority (Teaching)

Pitman

PITMAN BOOKS LIMITED
128 Long Acre, London WC2E 9AN

PITMAN PUBLISHING INC.
1020 Plain Street, Marshfield, Massachusetts

Associated Companies
Pitman Publishing Pty Ltd, Melbourne
Pitman Publishing New Zealand Ltd, Wellington
Copp Clark Pitman, Toronto

First Published 1976
Second Edition 1982

British Library Cataloguing in Publication Data

Anderson, John, 1921
The multiple choice question in medicine.
–2nd ed.
1. Medicine – Examinations
2. Grading and marking (Students)
3. Multiple-choice examinations
I. Title
610'.76 R834

ISBN 0-272-79642 5

Text set in 9/11pt Linotron 202 Plantin, printed and bound in Great Britain at The Pitman Press, Bath

The Multiple Choice Question in Medicine

Preface

Multiple choice questions are widely used in medical examinations but while they have now gained general acceptance as an assessment method there is still a hint of uncertainty and suspicion about the technique in some quarters. Sometimes this is because of ignorance and lack of experience, sometimes because the failings and faults of poor MCQ have been unduly emphasized, and sometimes because the examiners expect too much from them or try to use them for purposes that would be better served by some other technique. The fact remains, however, that carefully prepared MCQ are unrivalled as reliable and highly discriminatory objective tests of factual recall. At the same time higher areas of taxonomy can be assessed by means of well thought-out questions, without loss of discriminatory power.

I do not intend this to be a comprehensive theoretical review of the MCQ technique and its place in medical educational technology. The aim of the book is essentially practical, although theory is by no means ignored. The chapter headings describe the content and it is hoped that the book will be helpful both to examiners and candidates. Each group will probably concentrate on different areas, although I hope that there will be something of interest for everyone in each chapter – candidates may gain reassurance as well as information from some of the more theoretical sections.

The book is heavily biased towards the multiple true/false type of question and the Newcastle University computer program used to mark and analyse such questions. I offer no apologies for this bias – this type of question and the Newcastle program are probably used more widely in the UK than any other single format. Here it is appropriate to acknowledge the major role that Professor E. S. Page (now Vice-Chancellor of the University of Reading), Dr W. Anne Wood, Dr Geoffrey Shearing and their colleagues in the University of Newcastle Computing Laboratory have played in the evolution of MCQ techniques over the last fifteen years.

There may be some repetition in the text; this is because I have attempted to make each chapter an essay complete in itself, and have tried to avoid too many distracting cross-references. At the same time unnecessary repetition has been avoided by placing all the references mentioned in the text in a separate

bibliography (p. 85). This bibliography is not intended to be comprehensive, but refers to the more important, relevant and recent papers on the subject. The two hundred multidisciplinary questions at the end of the book are intended as an educational as well as an assessment exercise, and comments are given on nearly all items.

The success of the first edition of this book has encouraged me to be a little bolder in the second. The text has been completely revised and lengthened and I have included much more theoretical discussion, not only in relation to the MCQ technique alone but also to assessment methods in general. The statistical and mathematical aspects have been extended and although the emphasis is still fairly and squarely on the MCQ, a broader field has been explored and, at times, a more philosophical approach adopted.

Forty-one of the original one hundred and fifty questions have been replaced and most of the remainder revised and updated. In many cases comments on the answers have been expanded. Fifty new questions have been added, giving a total of two hundred and, I feel, providing a better balance and more comprehensive coverage.

If the book helps examiners to set better questions and to analyse examination results more critically, and if it assists and reassures candidates, while clarifying the technique for them, I shall be happy.

Acknowledgements

This book could not have been written without the co-operation and interest shown by Professor E. S. Page, previously Director of the University of Newcastle upon Tyne Computing Laboratory, and of Dr Anne Wood, who was responsible for a number of years for processing MCQ, writing appropriate programs and for seeing that the suggestions of examiners were translated from theoretical concepts into practical procedures. Miss E. D. Barraclough, Executive Director of NUMAC, has provided invaluable help over the years.

Since the publication of the first edition a succession of staff in the Newcastle Computing Laboratory have continued and even extended the invaluable advice and assistance offered by their progenitors. I would particularly like to acknowledge the help of Dr Geoffrey Shearing, Dr Valerie Sargent, Mrs Gail Barker, Mrs Ann Turner and Mrs Elinor Robertson. It has been a constant pleasure to work with these congenial and expert colleagues and in the few crises we have faced we have aged together.

I would like to acknowledge the continuing advice and encouragement of Professor G. P. McNichol, Dr Peter Fleming, Dr Peter Sanderson and Dr J. F. Stokes. Over the last ten years I have learnt from them (and from other colleagues) more about MCQ (and about medicine) than I believed possible.

Above all, I must thank Professor Sir George Smart who introduced me to MCQ many years ago and gave me the opportunity to involve myself in a fascinating and rewarding educational experience. I am grateful to Professor Sir John Walton, Dean of Medicine; Mr Norman Shott, lately Deputy Registrar; Mr Derek Nicholson, Senior Assistant Registrar; and the Board of the Faculty of Medicine at the University of Newcastle for allowing me to refer to the Newcastle examinations and giving me permission to publish questions used in these examinations. Although I edited these questions I did not compose them all myself and I am grateful to my many colleagues who were their authors, and who continue regularly to produce questions.

I would like to thank the secretaries who brought this book to fruition: Mrs Jacqueline Home, Mrs Bunty Matthewson, Mrs Ann Potts and Miss Joan Pattison (what is the collective word for a group of secretaries?). My wife and family have again tolerated me patiently while the second edition was being prepared, and my daughter has helped me with much of the more advanced mathematics. Finally I must thank the staff of Pitman Books for their unfailing patience and courtesy.

J. Anderson
Newcastle upon Tyne
1981

Definition of Terms

Confusion has arisen in the past because terms in common use in connection with MCQ mean different things to different authors. The definitions given here are those that the majority of postgraduate examining bodies, many medical schools and the University of Newcastle Computing Laboratory have found convenient.

1. A QUESTION is the basic component of a multiple choice paper. It comprises:
2. A STEM or introductory statement and a series of
3. ITEMS, identified by a letter. Items are the smallest component of the examination to which the candidate may make a response. The terms COMPLETION and ALTERNATIVE are synonymous with the term item.
4. The word STATEMENT is used at some points in this book. It really has nothing to do with MCQ and is used purely for convenience. An item may in itself be a complete statement or the stem plus an item may together make a statement. It is to this statement that the candidate responds.
5. A RESPONSE is the positive selection made by the candidate in marking the RESPONSE or ANSWER SHEET – whether the response be 'true', 'false' or 'do not know'. The response may be *correct* or *incorrect*, depending on whether it corresponds with the examiners' response or not. The candidate's SELECTION is in our vocabulary synonymous with his RESPONSE.
6. TRUE is *not* synonymous with CORRECT, or FALSE with INCORRECT. True and false refer to either the candidates' or the examiners' responses; by definition the examiners' responses will always be correct! The candidates' responses may be *correct* or *incorrect*.
7. The MEAN SCORE is the mean of all the candidates' scores for an individual question or for the paper as a whole.
8. 'PERCENTAGE CANDIDATES CORRECT' is exactly what it says and is given on the computer printout for each individual item. When

converted to a decimal (80% = 0·80) this figure is referred to as P – an index that indicates the degree of difficulty of an item. (See Chapter 5.)

9. An INDEX OF DISCRIMINATION may refer to a question or to an individual item and gives an indication of how well that question or item discriminates between the better and the weaker candidates. In the Newcastle Marking Program the index of discrimination is a correlation coefficient. (See Chapter 5.)
10. RELIABILITY is a measure of the consistency and precision with which an MCQ paper (or indeed any assessment method) tests what it is supposed to test. Independent assessments of a single personal attribute should, in theory, yield identical results. In the Newcastle Marking Program the internal reliability of an MCQ paper is assessed by using the Kuder–Richardson (KR) formula number 20, modified for +1, −1, 0 scoring. (See Chapter 5.)
11. VALIDITY is another criterion used in the assessment of multiple choice questions or papers. A valid test is one that measures 'what we want it to measure, all of what we want it to measure and nothing but what we want it to measure' (Thorndike and Hagen, 1969). Validity has been subdivided into CONTENT VALIDITY, CONSTRUCT VALIDITY and CRITERION-RELATED VALIDITY. (See Chapter 5.)

Contents

1. Types of Multiple Choice Questions

Examinations have been used since time immemorial to assess student performance, to determine whether certain standards have been achieved and to decide whether teaching objectives have been attained. One of the important functions of examinations is to rank students – the most able being graded highly and the less able being placed lower down the list. It is clearly important that such a ranked order should accurately reflect the candidates' abilities and that the grading should be as objective as possible – relying principally on performance in the examination and not depending on the foibles or idiosyncrasies of the examiners. An examination should be fair and free from examiner-bias, and the results should be influenced only by those factors that the examination is designed to test. A 'poor' examination, one that is ambiguous or one that depends more on a knowledge of the intricacies of the English language than on a knowledge of medicine, will produce results that must be interpreted with reservation.

Aims of Examinations

The traditional medical examination forms include essay questions, oral confrontation and clinical and practical tests. Each form tests something different, and the subjective nature, inconsistency of marking and lack of reproducibility of these examinations (particularly essay questions) are well recognized (Bull, 1956; McGuire, 1966; Fleming *et al.*, 1974; Smart, 1976). Nevertheless, each type still has a role to play in the assessment of undergraduates and postgraduates, always provided that the effect of examiner variability on the scores can be reduced to a minimum, and that each candidate is given a fair and equal chance of demonstrating his abilities. The objectives of any examination should be clearly defined – different forms may have different aims (see pages 5, 55–6). It would be a sad day for medicine (and for prospective patients) if the clinical examination were to be discontinued. The oral examination, properly conducted by experienced examiners, offers a unique opportunity for the candidate to demonstrate his particular strengths, and the

essay question, carefully set and assessed, allows the candidate to exhibit his ability to analyse and discuss a particular problem. The essay question that simply tests factual recall (still all too common) is as much a waste of candidate and examiner time as the oral that is used to test only factual knowledge.

Bloom and his colleagues (1956) divided educational objectives into three 'domains' – cognitive, psychomotor and affective (knowledge, skills and attitudes). A successful doctor must acquire ability in all three domains, and appropriate tests should be used during the training period (both undergraduate and postgraduate) to determine whether such ability has been attained. Charvat, McGuire and Parsons (1968) (quoted by Fleming *et al.*, 1976) have extended the original taxonomy of Bloom and his colleagues and have modified it specifically for medical examinations. In relation to the cognitive domain, Charvat and his colleagues define six levels (the CMP classification):

(a) Knowledge of fundamental facts, concepts, principles, laws, methods and procedures.
(b) Understanding of these facts, concepts, etc.
(c) Ability to understand and interpret data.
(d) Ability to solve relevant problems.
(e) Judgement in evaluating a total situation.
(f) Ability to create a new synthesis.

Levels (a), (b) and (c) can readily be tested using objective methods (see below); to some extent level (d) can also be tested in this way (Anderson, 1980b). The higher taxonomic levels in the cognitive domain, (e) and (f) cannot easily be assessed like this, and here the essay question and the oral examination, if properly used, have a part to play. The 'study in depth' (Pickering, 1978), as used in several medical schools, can test the whole range of the CMP classification, although the areas of knowledge studied and tested will usually be limited.

Tests in the psychomotor domain include 'practical exams' which test laboratory skills. These are used much less frequently than in the past, since it is now widely recognized that the ability to carry out laboratory tests manually and to manipulate equipment is much less important than an understanding of the principles and concepts underlying such procedures. There are much better ways of assessing these skills than letting a student loose in the laboratory, and when such assessment is used it now most commonly takes the form of data recognition and interpretation, thus moving the tests (rightly) into the cognitive rather than the psychomotor domain. The clinical examination (traditionally 'long-cases' and 'physical signs') is another method of assessment in the psychomotor field; here, clinical competence and the ability to communicate are tested.

Assessment in the affective domain is the most difficult of all. Charvat *et al.* (1968) have listed the important attributes in this area as:

(a) Acceptance of responsibility for patient welfare.
(b) Concern and consideration for patient and patient's family.
(c) Recognition of medical capabilities and limitations.
(d) Ability to establish effective relationships with colleagues and other members of the health team.
(e) Regular observation of appropriate safeguards.
(f) An inquiring mind.
(g) Willingness to use medical capabilities to contribute to community as well as individual patient welfare.

As Fleming *et al* (1976) point out, 'a requirement to test all candidates in these respects in a final examination would be greeted by most examiners with dismay, not to say derision'. However, these authors discuss current views and techniques concerned with the testing of professional attitudes in Medicine. Such assessment is most commonly (although only partially) achieved at present by the use of continuous or in-course assessment, where teachers study and evaluate a student's performance during a wide range of clinical attachments. Such assessment and evaluation continues during the pre-registration year.

From the foregoing discussion it will be appreciated that when examiners consider assessment of undergraduates or postgraduates they should constantly ask themselves firstly 'what particular attributes or skills are we trying to test?' and secondly 'what is the best, most appropriate and most effective way of testing these?'

Objective Methods of Assessment

During the last twenty or more years, medical educationists have spent much time in attempting to develop methods of assessment that have been more objective, valid and fair, while indicating accurately the degree of attainment by the candidates in the particular skills the examination is designed to test. Among the important attributes of an examination are that it should be reproducible and that all candidates should be treated and assessed equally. It is for these reasons that multiple choice questions (MCQ) have been adopted and are increasingly used as one form of objective assessment in medicine.

An objective examination is one in which each candidate answers the same paper under the same conditions, all responses being assessed in exactly the same way, without any subjective bias or influence from examiners. The mark

given for every possible response is determined beforehand. In addition to MCQ, objective tests that have been used in medicine include the 'erasive patient management problem' in which a case history is presented and a series of options offered, opposite each of which is an instruction, or additional information concealed from the candidate. Having made his choice, the candidate is able to reveal the item by using one of a variety of techniques to erase the over-print mask. In this way the candidate works through a series of decisions, and marking depends on the appropriate nature of his choices and the number of decisions he has taken in 'managing' the problem presented. The modified essay question (Knox, 1975) is another useful objective technique. Harden and Gleeson (1979) have described an Objective Structured Clinical Examination (OSCE) and this method of assessing clinical competence is effective, fair and accurate, albeit rather tiring for patients and sometimes boring for examiners. The OSCE is complex, must be organized with great care and may require complicated mathematical analysis to yield the results. Furthermore, the student is not tested on his ability to look at the patient as a whole. Nevertheless, this form of examination can be combined with the more traditional 'long-case' to give an overall assessment of a student's clinical competence and his ability to communicate. The technique is being used more and more widely, both as a formal and an informal assessment method. Fleming (1975) has described the missing-link question (MLQ), which is a method of testing candidates' ability to recall and manipulate factual information. This format comprises the insertion of a term which forms a logical link between two other, given, terms. This method seems to discriminate reliably between good and poor candidates.

The short answer (or short essay) question has been shown to have high inter-examiner reliability (Wakeford and Roberts, 1979; Anderson, 1980a); perhaps because the very nature of such questions is that fundamentally they test factual recall. By defining a 'check-list' of items of knowledge that candidates should be expected to include in their answers, this type of assessment can be made much more objective. Short answer questions (unlike MCQs) do not provide 'cues' for the candidates and a number of workers have pressed for the increased use of such questions in medical examinations (Forsdyke, 1978; Hettiaratchi, 1978; Wakeford and Roberts, 1979; Anderson, 1980a).

However, the most familiar objective examination, and certainly that most commonly employed, is the MCQ paper, which has been used for many years as a reliable measure of knowledge and as a means of accurately discriminating between candidates. Fleming *et al.* (1976) have emphasized that 'whatever the scoring system employed, an MCQ paper can do no more than place the candidates in rank order in terms of their ability to answer that particular test'

and therefore such papers should be valid tests of knowledge (see page 59) and should be set at a level of difficulty appropriate for the examination in which they are used. The MCQ paper mainly tests recall of factual information and has the unrivalled ability to give an unambiguous and reproducible assessment of knowledge over a wide field. It is, however, now accepted that good MCQ can test more than simple factual recall and a well-prepared paper can test powers of judgement and discrimination as well as the ability to reason. As Lennox (1974) has stated, the intelligent application of well-understood principles and probabilities will often earn better dividends than plain memory work in answering an MCQ test. It must nevertheless be appreciated that different types of examination test different strengths and weaknesses, and it would be unreasonable to expect a conventional MCQ paper to assess adequately all of those qualities that the qualifying undergraduate must possess. Other assessment methods, carefully conducted, are often more appropriate – it has never been claimed that MCQs test the candidate's ability to express himself concisely, lucidly and grammatically, and it is unreasonable to expect the technique to do this. They cannot readily be used to assess the highest taxonomic levels in the CMP classification, nor can they test, in their conventional form, in the psychomotor and affective areas. On the other hand it is equally important to remember that other assessment methods (essay papers, oral and clinical examinations, etc.) should *not* be used simply to test factual recall – MCQs do this much better and more reliably.

In the earlier days of MCQ, students and teachers often regarded the technique with suspicion, and some still do. Students protested against its impersonal nature (until they appreciated that it was essentially fair) and, justifiably, against the ambiguity of poor questions. Teachers were often uncertain what qualities were in fact being tested and sometimes felt that their roles as assessors were being usurped by the computer. This anxiety was due to a combination of unfamiliarity with the technique, suspicion of a new method of assessment in which the teachers were only involved in the setting of questions and not the marking of the answers, a high prevalence of poor questions that brought the technique into disrepute in some quarters, and finally the fact that some teachers expected far too much from the MCQ paper and did not realize its limitations. To some extent anxiety about the method persists today (Dudley, 1973; Pickering, 1979), but in general MCQs have now been accepted both by students and teachers as an important and valuable means of assessment. Just as acceptance has increased over the years, so the questions used have increased tremendously in 'quality' (with a consequent increase in the reproducibility, discriminating power and validity of the examinations) and methods of analysis of the results have become much more sophisticated, valid and useful. The MCQ has been the subject of many

scientific studies that have given valuable information covering the whole field of medical education (Anderson, 1979).

Among such studies reported in the last few years the following may be of particular interest both to examiners and candidates. Ricketts *et al.* (1974) published a comparison of performances of MB candidates from three Universities in MCQs in medicine and Stevens *et al.* (1977) carried out a similar study, comparing performance on MCQ papers in obstetrics and gynaecology at five London Medical Schools. Anderson *et al.* (1979) reported a comparison of performance in a multidisciplinary MCQ paper between Newcastle and Jordanian students. Gandi and Ghosh (1978) published a trial of MCQ examinations in pharmacology, and Fredman (1979) reported an analysis of behavioural objectives in the teaching of anatomy from the results of marks in MCQs. Skakun *et al.* (1979) described a preliminary investigation of three types of MCQ, and several papers have compared MCQ tests with other assessment methods – short answer and MCQs in the evaluation of students of biochemistry (Forsdyke, 1978), short answer and MCQs in physiology (Hettiaratchi, 1978) and multiple choice and free-response tests in examinations of clinical competence (Newble *et al.*, 1979). Studies have also been carried out on the prediction of student performance by multiple choice testing (Sanderson, 1973a; Anderson, 1978). This list of publications is by no means complete; it simply indicates the scope and breadth of studies based on MCQs.

Types of Multiple Choice Question

The simplest form of 'multiple' choice question is the yes/no alternative, e.g.

Is Edinburgh west of Newcastle?
A. Yes
B. No
(Correct answer: A)

This simply tests factual recall and nothing else; it has no place in medical examinations.

In the UK two basic formats of MCQ are used. Both consist of an initial statement (or stem) followed by a number of possible completions or subsidiary statements (items).

The one-from-five type This asks the student to select a single correct answer from a group of five alternatives. It is important to recognize that in this format what is wanted is not necessarily the *only* possible answer but the 'best' answer

of those given and it is wise to reinforce this principle by the wording of the stem. Four examples of the one-from-five type of question are shown below.

A 46-year-old male patient who is moderately obese is suspected of having Cushing's disease. The *single* observation which would give *most* support to this diagnosis would be

A. urinary 17-oxosteroid excretion of 80 μmol/24 h (23 mg/24 h)
B. a midnight plasma cortisol of 555 nmol/l (20 μg/100 ml)
C. a blood-pressure of 170/110
D. a diabetic glucose tolerance curve
E. the presence of pink abdominal striae

(Correct answer: B)

The most reliable *single* test that would best support a diagnosis of primary hypothyroidism in a 35-year-old female patient who is taking an oestrogen/progestogen oral contraceptive would be

A. estimation of serum thyroxine concentration
B. estimation of serum cholesterol concentration
C. examination of an electrocardiogram
D. examination of the blood for the presence of circulating antibodies to thyroid components
E. estimation of serum TSH concentration

(Correct answer: E)

A man of 68 presents with a sudden episode of severe colicky abdominal pain. Within three days of admission to hospital he becomes jaundiced and develops a high fluctuating pyrexia. The *most likely* diagnosis in this patient is

A. infective hepatitis
B. cirrhosis of the liver
C. a stone impacted in the common bile duct
D. carcinoma of the head of the pancreas
E. an intrahepatic abscess

(Correct answer: C)

In a patient with suspected osteitis deformans (Paget's disease of bone) which *one* of the following investigations would be of most value in confirming the diagnosis?

A. skull X-ray
B. measurement of serum calcium concentration