2. Modified essay question (MEQ) paper: perestroika

Cameron Lockie, BSc, FRCGP, Sean McAleer, BSc, DPhil., Helen Mulholland, MA, Ph.D
Roger Neighbour, MA, FRCGP, and Philip Tombleson, MBE, FRCGP

SUMMARY: Traditionally the modified essay question (MEQ) paper has attempted to test problem solving and decision making based on an on-going family saga using seven or eight questions to be answered in 90 minutes. Candidates' scripts are double marked by two College examiners. This format imposes constraints on the range of questions asked and results in contrived scenarios. It is possible to be 'coached' for this and double marking is expensive in examiner time.

Recent studies show that validity and reliability are improved by increasing the number and range of questions in a 'surgery type' paper. Single marking has been instituted and the MEQ paper will in future consist of 10 or more questions to be answered in 2 hours. Examiners' marking performance is monitored by senior examiners.

Technical and statistical considerations are discussed, as are implications for candidates and course organizers.

Introduction

The assessment of a doctor's attainments is so complex and varied an activity that general practice requires a correspondingly complex and varied examination. The MRCGP examination seeks to test candidates' knowledge base; the use they make of this knowledge in problem solving, decision making and management; their clinical and other professional skills; and their attitudes to the job, to their patients and to their own personal and professional development. Ideally, each component of the examination should test in the areas which it alone is good at, though in practice there is bound to be some overlap. Thus the MCQ paper has proved a highly discriminating and reliable test of candidates' knowledge base, yet to date it has not been found feasible to include a clinical component in the MRCGP examination, so candidates' clinical skills and attitudes have been tested inferentially in the MEQ paper and the two oral examinations. Traditionally, the MEQ paper has tried to test problem solving and decision making in a broad range of clinical settings rather than in great depth. However, the MRCGP examination is undergoing a reappraisal of its aims and methods, and the MEQ paper has entered a phase of transition, which is discussed in this paper.

History of the MEQ

Now widely used in undergraduate and postgraduate medical education in the United Kingdom and throughout the world, the MEQ paper in its present format is largely the result of development by the Royal College of General Practitioners (Feletti and Smith, 1986; Rabinowitz, 1987). It has two principal uses: as a formal assessment technique; and as an aid to learning, by supplementing lectures, group discussions, and programmes of self-audit and self-instruction (Knox, 1975).

The 'traditional' MEQ paper introduces a central patient and family, who are presented question by question in a saga of medical problems whose gradual unfolding over time mimics the uncertainties and time-scale typical of general practice. Early MEQ papers had 10 questions, but in recent years there has been a tendency for this to drop to eight or even seven questions. The case history provides pegs on which to hang questions about the recall and interpretation of knowledge, medical priorities and management, problem definition and solving, the psychological, family and sociological aspects of illness, doctor-patient communication, and (to some extent) doctor attitudes.

A search through the last decade's MEQ papers would reveal a number of recurrent themes. These include:

- clinical medicine, including chronic illness, prescribing, and preventive medicine
- 'problem' patients, whose personalities, expectations, circumstances and lifestyles may affect management
- psychological problems of individuals and families, and their treatment
- the consultation process and dynamics
- general practice organization, including the primary health care team and problems of practice administration
- relationships with other medical and paramedical colleagues
- controversial and 'hot' topics, e.g. alcohol, AIDS, alternative medicine, environmental medicine
• the doctor's own feelings, stresses and self-awareness
• ethical and attitudinal issues.

The examiners responsible for setting the MEQ paper have for several years been aware that for trainees, who now form the majority of those taking the examination, the traditional case developing over many months or even years has not been representative of their everyday experience. Accordingly, recent papers have tended to be of a newer 'surgery' type, where the clinical scenario is of a single surgery or a single working day during which a succession of problems is presented. Although something of the majesty of general practice timescales is lost in this format, it is generally felt to be fairer to young doctors with a limited experience of practice.

Reviews of the examination

In February 1987, at the request of the then Membership Division of the College, Dr David Swanson of the American Board of Internal Medicine produced a review of the membership examination procedures (Swanson, 1987). His report, while in general highly complimentary to the examination, highlighted a number of areas of potential improvement. A preliminary analysis of the MEQ paper had indicated that a candidate's performance might vary considerably from one question to the next, irrespective of differences between markers. In order to increase the reliability of the paper, Swanson suggested that the number of questions should be increased and that the questions should as far as possible be independent of each other in context and topic. He foresaw that this might require an increase in the duration of the MEQ paper beyond the usual 90 minutes.

A further, more detailed study of the MRCGP examination was begun in 1988 by Dr Helen Mulholland and Dr Sean McAleer of the Centre for Medical Education, University of Dundee. Their work is scheduled to last for three years, but has already resulted in a number of changes in the MEQ paper (Mulholland and McAleer, 1988).

Mulholland and McAleer have recommended that, in order to improve the reliability and validity of the MEQ paper, the number of questions be increased to at least 10 and if possible 12 (see page 21). Pre-empting a protest by the MEQ examiners about increased workload, they also established that it would be possible for the MEQ paper to be 'single marked', that is, each question could be marked by one examiner, instead of the previous two, without any significant loss in reliability (see technical and statistical considerations below). This procedure was implemented for the first time in the May 1989 examination.

The Dundee researchers reminded the examiners of the need for rigorously devised marking schedules which should be scrupulously adhered to. The examinations are currently endeavouring to put this sound advice better into practice. Another suggestion, as yet not taken up, is the concept of 'isomorphism', whereby an 'item bank' of questions could be developed using the same scoring schedule but having differing surface structure. This may increase reliability, but possibly at the expense of some contamination and leakage!

Isomorphism is explained by the use of MEQ sample question 7 on page 51.

The surface features of this problem are the age and sex of the patient, the actual condition (sarcoma), the number and ages of their children and the mother's occupation.

The deep structures consist of the following areas of involvement:

Family dynamics: How are the other children likely to act?

Primary health care team: What other members of the team will make a contribution to the care of this family?

Doctor's feelings: How does one handle the death of children and the emotional interactions within the family?

Insight: What are the possible areas which may present particular problems in this situation?

Management: How may palliative care best be provided for this patient?

These deep structures could be used to write a scenario based on a girl aged, say 14, suffering from a brain tumour. The girl has two sisters aged 13 and 18 and is to be cared for at home. Her mother is a practice manager.

The surface structure of this question could possibly be changed even more, for example by introducing the death of the mother, but still test the same areas: or altered yet further by taking the idea of testing the areas listed, using a family of any size. The primary health care team could be involved with reference to only the last three areas related to any terminal case (doctor's feelings, insight, management).

The example given in question 7 samples one case only from all possible terminal cases at a surface level, but concerns the underlying attributes which are involved in all such situations. One is not merely testing the ability to deal with this situation in particular, but using it as a device to obtain information about the handling of all similar problems. This is termed 'generalizing' (from the particular case to the general type of situation) and forms the basis of generalizability theory.

Reasons for change

Preceding paragraphs have indicated some of the historical and theoretical background to the present changes in the MEQ paper. The traditional MEQ paper imposed some constraints on the range of questions that could be generated from a single case, and the succession of way-out and contrived scenarios has sometimes strained credulity. The new 'surgery' format seems to be an improvement in this respect and can now be expected to become the norm.

However, the feeling has grown amongst examiners (and medical educators) that the observed gradual rise
in candidates’ mean scores in the MEQ paper may have more to do with learning the techniques of answering it than with any genuine rise in intellectual or intuitive standards. Some candidates attribute to the proverbial bus load of health visitors and so on a miraculous ability to solve every problem and score maximum marks. Every diet of the examination brings its crop of ‘course clones’ who answer every question under the headings ‘physical’, ‘psychological’ and ‘social’ factors even where such artificial divisions are patently inappropriate. The shallowness of candidates who merely pay lip service to such notions is readily detected in the orals, and one wonders whether their real interests have been well served by ‘teaching formula’.

So, if the job of those who run MRCGP courses is to prepare trainees for the examination, it is the job of the panel of examiners to keep at least one jump ahead of them. In this sphere of medicine, as in most others, to stand still is to fall behind, and it is the aim of the MEQ paper to become less ‘coachable’ as well as more valid, more reliable and more discriminating. The next section reviews some of the changes already implemented.

Changes and new ideas

Length of paper

In future the MEQ paper will consist of a minimum of 10 questions. (Two-part questions have fallen into disfavour.) All questions will carry the same number of marks. The time allowed for the paper has been increased from one-and-a-half to two hours.

Marking

The setting of the MEQ paper is the responsibility of a small national group, who meet at intervals to generate and refine the questions. Other members of the panel assist by suggesting questions, and by answering successive drafts under examination conditions. Marking schedules are devised at the annual examiners’ workshop. Until May 1989, cells of five or six examiners were responsible for marking individual questions. Papers were ‘double marked’, with the second examiner being unaware of the marks awarded by the first, and a composite mark awarded. Since May 1989, each question has been ‘single marked’ by one examiner. In order to maintain consistency, a senior and reliable examiner is appointed ‘marking monitor’ for each question. After the final examination results have been announced, the monitor re-marks a randomized sample of scripts marked by other examiners in the marking cell, the marks are analysed statistically, and are then fed back. The current target is for 10% of scripts marked by new examiners and 5% of those marked by established examiners to be monitored.

Extending the curriculum

All components of the MRCGP examination unashamedly claim the right, by asking questions in new forms or on fresh topics, to extend the curriculum into areas the College considers relevant or desirable. Recent examples include questions on chlamydia, AIDS, cardiopulmonary resuscitation, clinical protocols and current ethical dilemmas.

New types of question

Recent MEQ papers have included supplementary material such as a problem summary card or a protocol for a ‘well man clinic’ on which questions can be based. This trend is likely to continue. Also, as the PTQ paper is shortly expected to undergo radical change, it is likely that the MEQ paper might subsume some of the PTQ paper’s previous territory, such as examining a topic in depth, or testing the ability to write coherent prose (as in a letter of referral). Innovations such as these will offset the ‘coachability’ referred to earlier.

Some sample questions and ‘model’ answers are given in Appendix 2.

Technical and statistical considerations

Those who create and mark examinations sometimes behave as if the differences between the scores were caused solely by the individual differences between candidates. They assume that someone who scores 90 is automatically ‘better’ than someone who scored 85. In fact not all the variance in scores is candidate variance. In the MEQ paper differences between scores may also be caused by:

- differences between questions
- differences between markers
- mistakes.

Technical jargon would describe these as question, marker and error variance, variance being a measure of the variability of the score.

Question variance

Question variance is easily understood. Not all examination candidates are equally well informed about all the areas of general practice. If a large number of questions are asked, candidates are most likely to be tested in a range of both well known and unfamiliar areas. Question variance is a major factor in the MEQ paper.

Generalizability theory (Brennan, 1983) is a technique that extrapolates from existing data to predict the reliability which would be achieved by changing the structure of the examination paper. The application of generalizability theory indicated that by increasing the number of questions to 10 the reliability of the paper could be greatly enhanced. The measure of reliability used is the alpha coefficient (Cronbach and Gleson, 1972). This is basically a measure of the degree of consistency within a test. The higher the value the greater the reliability. An acceptable value is 0.80.

Marker variance

By applying this technique Mulholland and McAleer examined the effect of marker variance. Using two
markers does not double the reliability of the examination; in fact it raises the alpha coefficient only slightly. It is more cost-effective to have two questions single marked than one question double marked. In Table 1 sections A and B show the effect of increasing the number of questions while C shows the effect of increasing the number of markers.

Error variance

The third source of variance – errors of one type or another – is already minimized by the careful procedures for transcribing and computerizing marks. The single marker method was compared to the double marker method using a statistic called the 'standard error'. This statistic reflects the range of scores within which a candidate's actual score must fall. It is the estimate of the size of error. The standard error on the MEQ paper with seven questions and two markers was of the order of 5%. With 12 questions and one marker this should fall to 4%.

Subjectivity

In any assessment marks should be awarded for the presence of certain specific features in the candidate's performance. If these features are not present marks should be withheld. In essay marking, marks may be awarded for accuracy and amount of information, logical organization of content, literary style and/or orthography (handwriting and spelling). Mulholland and McAleer analysed a number of MEQ answers but could find no significant correlation between their necessarily subjective assessment of style and marks awarded by examiners. Neither did poor handwriting or spelling appear to have an adverse effect on marks.

A preliminary analysis suggested that there was some relationship between the logical organization of content and the marks awarded, but since most essays were relatively well organized, this effect was not significant. Candidates usually use a system of paragraphing which reflects a logical approach to the question, although it may not directly relate to the marking schedule. We suggest that the main correlate of marks in modified essay questions is the amount of information in the essay.

Propositional analyses

The usual term for a 'piece' of information is a proposition. Propositional analyses confirm that high-scoring essays generally contain both more numerous and more relevant propositions than low-scoring essays. We counted the number of propositions which were directly relevant to the scoring schedule, column 1, and the total number of propositions, column 2. The proportion of the total number which was relevant is shown in column 3. We studied 50 low-scoring (17 out of 45) and 50 high-scoring (23 out of 45) MEQs and took the average. The samples came from seven different MEQ questions in the summer 1988 examination (700 essays analysed). High-scoring essays did contain more propositions and more relevant propositions than low-scoring essays (Table 2).

<table>
<thead>
<tr>
<th>Table 1. Generalizability coefficients – MEQ paper.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of questions</strong></td>
</tr>
<tr>
<td>---------------------------</td>
</tr>
<tr>
<td><strong>Number</strong></td>
</tr>
<tr>
<td><strong>Section A</strong></td>
</tr>
<tr>
<td>increasing</td>
</tr>
<tr>
<td>the number</td>
</tr>
<tr>
<td>of questions</td>
</tr>
<tr>
<td>1 marker</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td><strong>Section B</strong></td>
</tr>
<tr>
<td>increasing</td>
</tr>
<tr>
<td>the number</td>
</tr>
<tr>
<td>of questions</td>
</tr>
<tr>
<td>2 markers</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td><strong>Section C</strong></td>
</tr>
<tr>
<td>increasing</td>
</tr>
<tr>
<td>the number</td>
</tr>
<tr>
<td>of markers</td>
</tr>
<tr>
<td>7 questions</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Propositional analysis.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of propositions</strong></td>
</tr>
<tr>
<td>Relevant</td>
</tr>
<tr>
<td>High scores</td>
</tr>
<tr>
<td>Low scores</td>
</tr>
</tbody>
</table>

Implications for candidates

Although its content may alter, the format of the MEQ paper seems likely to continue without substantial modification. Members of the nuclear group who set the papers are confident that they can continue to devise MEQs which will reliably sort out the genuinely competent from the merely well rehearsed.

Specifically, MEQ papers of the future may seek to test more reliably in such areas as candidates' sensitivity and empathy, their interpretation and critical appraisal of clinical data, their ability to handle the unexpected, and their written communication skills.

There has been recent debate, notably within the Association of Course Organizers, about whether trainees risk becoming preoccupied with passing the MRCGP examination at the expense of their application to learning on vocational training schemes. Many examiners are fully experienced in both spheres. The MEQ paper tests candidates' ability to think laterally and logically, critically and comprehensively. We believe that these qualities, far from being at odds with
the aims of vocational training, are fully congruent with them, and that candidates who are enthused by the diversity of ideas current in vocational training are likely to acquit themselves well in the MEQ. As Descartes reminds us, "It is not enough to have a good mind. The thing is to use it well."

Acknowledgements

Particular thanks are due to the members of the MEQ paper nuclear group for their support, hard work and good humour: Roger Neighbour, Deputy Convenor (Convenor from October 1990), John Carroll, Graham Hornett, and Andrew Wilson. We would also like to thank Miss Sarah Baker, Practice Administrator, Rother House Medical Centre, Stratford upon Avon, for her excellent secretarial help.

References