

EXAMINING MEDICAL STUDENTS — THE MULTIPLE-CHOICE QUESTION PAPER

C. S. JONES, M.B., CH.B. (CAPE TOWN), *Department of Anaesthesia, University of Cape Town*

In 1956 Prof. G. M. Bull,¹ in reporting on his analysis of the final examination in medicine at Queen's University, Belfast, noted that examiner-error did not disappear from written papers until the multiple-choice question was introduced. Professor Elliot² has noted the same superiority of objective over subjective questions. In the objective multiple-choice question paper the examinee is expected to answer a large number of short questions and he is usually offered several answers, his sole task being to select the correct answer and to indicate his selection on a suitable sheet of paper. The answer sheet can be marked by anyone who possesses a key and can even be marked by machine. The questions can be framed in such a way that the examinee is tested on his factual knowledge or on his deductive ability. Variations in marks arise solely from differences between students in their ability to answer correctly the questions set.

There are several additional advantages noted by Professor Bull. One advantage is that it becomes possible to examine over a much greater field than can be covered in the conventional 3-hour, 4-essay type of written examination paper. Another advantage is that related fields can also be touched upon. In this way the examination will test not only the student's ability to retain pertinent facts from other subjects but also the effectiveness of the teaching in these subjects. This aspect is accentuated when the multiple-choice question is used in examining a large number of candidates from several different schools, since the examination will enable the standards within schools to be compared, as well as the standards achieved by individual students within a group or by groups of students within a school.

The present paper reports on the use of multiple-choice questions for comparisons within a class and for comparisons between classes within this medical school.

PRESENT STUDY

During the years 1953-1956 the class examination in anaesthetics was conducted using multiple-choice questions and, with minor exceptions which will be detailed, all students answered precisely the same set of questions, set in the same order and marked with the same key. Needless to say, the question papers were recovered at the time that the answer sheets were collected, and were not released for general study. While there was a loss of 4 papers in the very first year, there was never any evidence that this benefited subsequent candidates, who never knew in advance that they would receive the same set of questions as their predecessors.

The Examination

In 1953 60 questions only were set and 120 minutes were allowed for answering them, as this form of exam-

ination was entirely unknown to the students. Out of a class of 108 students, 1 completed the questions in 50 minutes, while at the end of the 120 minutes 4 had not yet handed in their papers.

In 1954 the number of questions was increased to 120 (the original 60 set in 1953 plus an additional 60) and

TABLE I. PERFORMANCE OF STUDENTS IN FOUR CLASSES, ANSWERING MULTIPLE-CHOICE QUESTION PAPERS

Total marks as %	Number of students				Total
	1953	1954	1955	1956	
96-100	1	—	—	—	1
91-95	1	—	—	—	1
86-90	11	1	—	2	14
81-85	16	5	1	4	26
76-80	16	11	5	13	45
71-75	25	18	15	13	71
66-70	20	21	18	19	78
61-65	15	19	22	11	67
56-60	2	11	24	10	47
51-55	—	6	6	5	17
46-50	1	1	7	3	12
41-45	—	1	—	2	3
36-40	—	—	—	—	—
Total	108	94	98	82	382
Mean mark and SD	75 (=9)	68 (=8)	64 (=8)	67 (=9)	67 (=10)

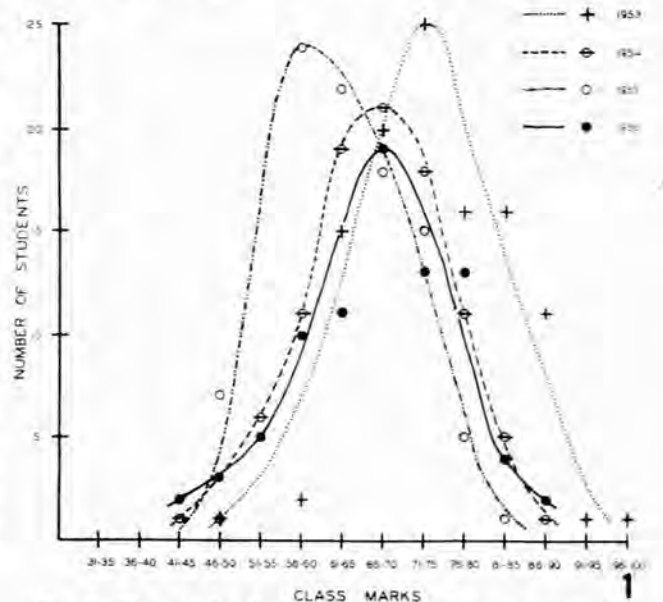


Fig. 1. The results detailed in Table I are graphically set out here to show the variations in the distribution curves from year to year.

180 minutes were allowed for their completion. In 1955, and again in 1956, the same 120 questions were used but only 120 minutes were allowed for completing the examination.

The performance of each class is set out in Table I and is also shown in Fig. 1. It will be seen that the time allotted for answering questions undoubtedly has some effect upon the efficiency of the class.

FACTORS AFFECTING CLASS PERFORMANCE

It is too often assumed that poor examination results indicate poor student material. Such an assumption is often very far from the truth and teachers are well aware that the material presented to the student, as well as the way in which it is presented, are factors which have an important bearing on examination results. Examinations test not only the ability of the students, but also the ability of their teachers.

There are a number of other variables which may affect the learning process, including such obvious ones as the health of the student and his natural interest in the subject being taught. In turn, there are variable factors which will affect the ability of the student to respond adequately to tests of his learning. The answers to some of the questions set in the anaesthetic class examinations show that whole classes may differ in their knowledge of commonplace matters.

Influence of Teaching

For the most part the response from year to year to the multiple-choice questions did not alter significantly but for some questions there was a progressive improvement suggesting that the subject was being more adequately taught. Thus, question 30 stated: 'The number of calories required to raise the temperature of one gram of substance through one degree Centigrade is a measure of its (a) latent heat of fusion, (b) melting point, (c) latent heat of vapourization, (d) specific heat, (e) respiratory quotient'.

The candidate had only to select the correct answer and ring it. In 1953, 12% failed to answer correctly but by 1955 only 3% failed.

Duration of Examination

Sometimes an increasing difficulty in scoring the correct answer suggested that the time available for answering each question (2 minutes in 1953 but only 1 minute by 1955) affected the result significantly, although it was not possible to exclude a progressive alteration in teaching. Thus, question 7 asked: 'The trigeminal nerve is primarily a sensory nerve. Which one of the following of its branches carries motor fibres? (a) Ophthalmic branch, (b) maxillary branch, (c) mandibular branch, (d) nasociliary branch, (e) mental branch'. Of the 1953 class 19% failed to select the correct answer, 45% failed in 1954, while by 1955 the failure rate had crept up to 60%.

Class Ability

In other questions there was a fluctuation in response which appeared to reflect the ability of the class more faithfully. In question 1 (quoted below) it was possible to discard 3 answers with some certainty on grounds of common sense alone. The student was left with 2 reasonably possible answers, (a) or (c), and the selection of the correct answer was still not particularly difficult. The

class of 1953 was the only one in which the historical fact here tested, was deliberately accented in teaching and their failure rate was 18%, but the class of 1954 (failure rate 30%) did not measure up to their colleagues of 1955 (failure rate 21%), who had less time available in which to deduce the answer. Question 1 read: 'The introduction of anaesthesia and the development of Lister's principles of antiseptics are perhaps the two greatest contributions to the art of surgery.' In connection with this quotation which one of the following statements is correct? (a) anaesthesia antedated Listerism, (b) anaesthesia and Listerism were perfected in different centuries, (c) anaesthesia and Listerism were perfected in the same year, (d) anaesthesia postdated Listerism, (e) the work of Lister made the development of anaesthesia possible'.

Chance

Since each question was provided with 5 possible answers, one of which had to be correct, it was possible to score 20% by chance alone, and to demonstrate this 20 people were asked to complete answer sheets at random and without access to the question sheets. Their answer sheets were then marked from the keys. The mean marks for these chance answers were 22% (standard deviation 4) and 21% (SD 5) for each of the 2 sections of 60 answers.

This fact, that where 5 alternatives are offered as answers for each question 20% can be scored by guesswork, was one reason for having each class set its own pass mark by calculating the mean mark and the standard deviation of the mean. But while 20% could be scored by chance, the odds on scoring 100% by chance alone were 1 in 5^{60} for a 60 question paper and much more astronomical for a 120 question examination. Chance, therefore, played a very minor role in the end, for each candidate had exactly the same chance.

Misinformation

It is obvious that question 67 was answered incorrectly (mean failure rate 93%) largely because the classes were misinformed, for if the whole class had guessed at the answer at least 20% should have guessed correctly. Question 67 was: 'The main factor determining the filling of the coronary vessels is: (a) the systolic pressure, (b) the pulse pressure, (c) the diastolic pressure, (d) the mean aortic pressure, (e) the pulse rate'.

Teaching

As already observed, it is not alone *what* a student is taught, but how he is taught, that is critical. To realize this, examine the response to question 35, which read: 'The percentage of oxygen in ordinary room air is: (a) 12.7%, (b) 14.6%, (c) 18.8%, (d) 20.9%, (e) 21.1%'.

The answer to this question is usually taught first in preparatory-school science courses and is certainly reiterated at the university level in both physics and physiology. Yet 74% of the 108 fifth year medical students in the 1953 class were unable to answer this correctly. Why there should have been so much difficulty in answering an elementary general knowledge question is in itself a question posing fascinating fields for speculation.

GRADING STUDENTS

There is little doubt that, in general, once a medical student has passed into the clinical years of study, it will

only be a matter of time before he graduates into the ranks of the profession. It is common experience that selection on academic ability is exercised largely in the first year of study (biology, physics and chemistry), to a lesser extent in the second year (anatomy and physiology) and to an even smaller extent in the third year (pathology, bacteriology and pharmacology). But if examiner-error plays a large part in determining the results of the conventional written examinations, this selection of students cannot be regarded as being naturally related to the ability of the students themselves. It becomes even less natural in the clinical years, when failure to pass the final M.B. examination casts, however obliquely, some shadow upon the teaching ability of the clinicians to whom the student has been apprenticed, and the natural bias of the teachers, who are in the main the examiners also, will be towards conversion of any shadow into reflected light. Most medical schools probably allocate marks somewhat on the lines of Queen's University, Belfast,¹ where 50% of the marks in the final examination in medicine are awarded on clinical and oral examination, the value depending entirely upon the examiner's impressions of the student's ability.

There is, in most walks of life, a form of natural selection based largely upon ability, but such ability is measured, not against those senior or junior to the person being measured, but against his contemporaries. It does not seem unreasonable to ask that such a natural selection of student material should also operate throughout the whole period of training and the use of the multiple-choice examination offers this. It is not necessary in such an examination to set an arbitrary pass mark and then resort to examiner bias or mathematical manipulations in order to ensure that an adequate proportion of the class leaps the hurdle. The calculation of the mean, and the standard deviation of the mean, of the marks scored by the participants in the examination will enable the examiner to reject about 17% whose marks fall below the mark determined by the calculation of the value of 'mean mark minus 1 standard deviation value'. These 17% are not rejected because they cannot achieve a standard set by the examiner but because they cannot achieve the standard set by their fellow students.

Reference to Table I shows that in this particular series of examinations, with a statutory pass mark of 50%, only 4% of the students will fail. But if the pass mark is adjusted to 70% to compensate for the 20% which can be scored by chance alone, then only 41% will pass. Thus more than half the class (55%) will fall within the range where chance alone, rather than ability, will determine their fate.

Where the class sets its own pass-mark one of the most prominent criticisms of the multiple-choice question paper falls away. Correspondence following the publication of Professor Bull's report leaned heavily upon the possibility that more than one of the answers provided might fit the question, according to the opinion of the critic or the examinee. But if the whole class is of the opinion that (a) is the correct answer when the examiner requires (b), then no student will be penalized at all. Those students who favour (a) when the majority of the class, as well

as the examiner, recognize (b) as the answer are, it is true, penalized for deviating from the popular view. This is a penalty which all original thinkers must learn to accept with dignity. The history of our profession is illuminated by numerous examples of physicians whose contributions were not recognized until after their death — a death not infrequently hastened by the scorn of their contemporaries.

To acquire an equanimity to such injustice at an early age will inevitably raise the standard of medical practice, as will a 17% loss from the student body in each of the 6 years of study, for only 35% of the students entering the first year will finally graduate where there is a loss of this magnitude in each of the 6 years. Actually, the use of the frequency distribution curve to determine which students are to pass an examination will not lead to a steady 17% rejection each year, or for each subject, for if the curve is positively skewed the class will have either a generally lowered ability in the subject in which they are being examined, or a poorer teacher, or both, and a larger percentage will tend to fail. The converse is of course also true.

If it is felt that the formula leads to too great a loss, then it can be adjusted downwards by multiplying the standard deviation value. Multiplying by 2 will lead to a failure of only 3% of the class.

SUMMARY

A discussion of the objective examination of medical students by means of the multiple-choice question paper is presented. With this technique it is possible to compare the general ability of succeeding classes of students and, by detailed analysis of the answers to each question, it is also possible to get some idea of the ability of each class in various subjects in the medical curriculum, as well as the ability of their teachers.

The results of the use of this type of examination technique are superior to more conventional methods since the individual student fails or passes a standard set by his colleagues, and not one set by an examiner who may be biased, quite unconsciously, when making border-line decisions.

While it may be quite true that a doctor should have a suitable personality, as Professor Bull suggests, it must be remembered that patients themselves differ in their personalities and it will never be possible to isolate and foster a 'universal' medical personality which will appeal to all patients. For this reason alone it is probably safer to exclude personality in grading medical students, and to concentrate rather on their ability to answer questions correctly. For, as he seeks medical aid, each patient asks, silently or volubly, that his medical attendant should provide factual details of diagnosis and prognosis. The doctor who can give the correct answer is the doctor the patient would like to have and since the medical profession exists only in order to serve the patient, perhaps we should use the patients' needs as a yardstick for medical training.

REFERENCES

1. Bull, G. M. (1956): *Lancet*, 2, 368.
2. Elliott, G. A. (1960): *S. Afr. Med. J.*, 34, 333.