

## The final examination in medicine: time for change?

J. M. Cacciottolo\*, A. Caruana Galizia\*

**ABSTRACT:** Background: An essential component of a programme of studies is its own evaluation. Most programmes culminate in a final assessment, in order that participants may be tested and graded. In June 1995, at the University of Malta, a group of 53 medical students sat for their final examination; medicine was one of the three co-equal component subjects of this statutory qualifying examination.

**Aims:** The scope of this paper is to analyze the results obtained in the final examination in medicine and to use this data to address such issues as aims of this examination, method and quality assurance of assessment.

**Method:** The result obtained by candidates in the final examination in medicine was correlated with their university entry qualifications. The composite mark for each candidate was split into its components and analysis took the form of description, correlation and clustering. Computation of Cronbach's alpha facilitated analysis of reliability of each of the three parts of the examination.

**Results:** Performance in the final examination in medicine was as follows: above average (grade B) - 3 (5.7%); average (grade C) - 31 (58.5%); below average (grade D) - 14 (26.4%). Five candidates (9.4%) failed the examination. Grade A was not awarded. The mean marks were: paper I (essays) - 12.9/20 (SD1.7); paper II (multiple choice questions (MCQ)) - 10.3/20 (SD1.8); clinical examination - 37.3/60 (SD4.3). The overall mean mark was 60.5% (SD6.4). There was only a weak correlation between these final results and university entry qualifications. Performance in the MCQ paper correlated better than the essay paper both with the clinical component and with the total mark. Of the examination components, the clinical part had the highest item total correlation (0.91), followed by the MCQ paper (0.73) and essays (0.64). As a measure of reliability, Cronbach's alpha indicates that the MCQ paper was most reliable and that the clinical examination was slightly more reliable than the essay paper.

**Conclusions:** The findings of this paper suggest that there is room for improving the quality of assessment methods. A review of methods and procedures, with the dual purpose of decreasing bias and increasing specificity and sensitivity of this statutory examination will not only benefit candidates, but ultimately also the University of Malta. The final qualifying examination in medicine should have clearly defined objectives and methods of assessment should be aimed specifically at reaching them. It needs to be able to assess the ability to think critically about diagnosis and management and to ensure that the candidate has a satisfactory base of factual knowledge. It also needs to assess objectively the adequacy of basic clinical skills and candidates' facility of communication.

\*Faculty of Medicine & Surgery, University of Malta, Medical School, G'Mangia

*Correspondence:* J. M. Cacciottolo, Faculty of Medicine & Surgery, University of Malta Medical School, G'Mangia

*Keywords:* evaluation, examinations, undergraduate, medicine

*This article is based on a paper read at the III Maltese Medical School Conference, 29th Nov-2nd Dec 1995.*

### Introduction

#### 1. Background & chronology

An essential component of a programme of studies is its evaluation. Evaluative procedures furnish an index of both pertinence and efficacy of the learning experience. Formative evaluation is used as a means of improving and developing a programme while summative evaluation is used for selection, grading or accountability; both types of evaluation are often construed of facets of the same procedure. The history of programme evaluation dates back to at least 2000 BC, when Chinese officials were conducting civil service examinations. The first formal educational programme

evaluation appears to have been conducted in 1897-1898 in the USA<sup>1</sup>.

The first recorded attempt at evaluating medical teaching in Malta was made by medical students in autumn 1986, albeit in the form of a lampoon. Then, a student rag magazine, "Apoplexy Now", produced a group calling itself Students' Organization for Reform of Medical Instruction, contained a list of 34 lecturers; each graded under eleven parameters together with satirical comment.

A scientific attempt at evaluating medical teaching was made by medical students in 1994 using an anonymous questionnaire as a tool for study<sup>2</sup>. The results created a certain degree of controversy and were not published. With regard to medicine (as distinct from

the other major specialties), the teaching of seven subspecialties making up the curriculum was assessed. There was little variation from an unspecified mean in the perceived quality of teaching of these subjects. In 1995, medical instruction for the clinical years was evaluated again by medical students, through a more comprehensive anonymous questionnaire<sup>3</sup>. Information was sought about teachers and teaching in pathology, medicine and surgery (but not obstetrics and gynaecology) using a four-point scale.

An opportunity to evaluate the medicine curriculum and examination methods was taken in mid-1995, when a group of medical students completed a five-year course of studies and sat for their final examination. Medicine was one of the three co-equal component subjects of this statutory qualifying examination, and the examining board for medicine consisted of fourteen members, including two external examiners. Both authors of this paper were involved in examining this particular student cohort.

## 2. Scope and purpose

This paper forms part of a larger study designed to evaluate the medicine curriculum and teaching methods. The aim of this paper is to analyze the results obtained in the final qualifying examination in medicine at the University of Malta Medical School. The purpose of this study is not only that of an academic exercise, but also addresses such practical issues as assessment methods, validity and quality assurance of examinations.

## Method

In the final qualifying examination in medicine, marks obtained by medical students as well as information about their university entry qualifications were obtained from the Office of the Dean of the Faculty of Medicine and Surgery. Entry qualifications were converted into a uniform numerical score in order to enable accurate inter-cohort ranking and statistical comparison with other data sets. As an example, a student who obtained an A (advanced) level grade A pass in each of three science subjects at one sitting had a score of 12.

The final qualifying examination results were dated 15.06.1995, and the overall grade reflected a score made up of marks from three components. The written examination consisted of four essays (paper 1); each question was marked separately by at least two examiners. Paper 2 consisted of sixty multiple choice questions (MCQ) with positive and negative marking determining the mark in this particular paper. For the three-part clinical/practical examination, each candidate was assessed separately by three teams of two examiners each; for this part of the examination, the overall mark was arrived at consensually by the six examiners.

Data deriving from examination results was analyzed using BMDP/Dynamic, Version PC90 (BMDP Statistical Software Inc., Los Angeles, CA 90025, USA). Analysis took the form of description, correlation, clustering and the computation of Cronback's alpha. The latter was used as a measure of reliability of the examination components and was computed from correlations between the factors making up the final mark as well as the A-level results.

## Results

### 1. University entry qualifications of the cohort and their relationship to the mark obtained in medicine in the final qualifying examination.

The actual entry qualifications of the fifty-three students who sat for the 1995 qualifying examination were converted to a mathematical score with integers ranging from 9 to 16. The median score was 13, with a mean of 12.7 (SD1.38). Twenty-eight candidates had obtained A-level passes in physics, biology and chemistry, all at grade A in one session.

Performance in the final examination was as follows; above average (grade B): 3(5.7%), average (grade C): 31(58.5%), below average (grade D): 14(26.4%), fail: 5(9.4%). Grade A was not awarded. There was weak positive association between the entry qualification score and the composite mark obtained in the final examination in medicine; this relationship, however, failed to reach statistical significance (Fig. 1).

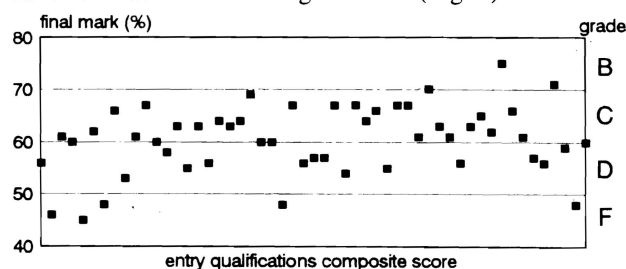


Fig. 1 - Relationship between the total mark obtained by candidates in the final examination in medicine and score of their University entry qualifications ( $r=0.27$  NS)

### 2. Analysis of the final examination results.

The overall mean mark obtained in medicine was 60.5% (SD 6.4) with a range from 45 to 75%. The written component of the examination, comprising two papers was each marked out of 20. The mean mark obtained in paper 1 (essays) was 12.9 (SD 1.7) with a range from 9 to 19. In paper II (MCQ), the mean mark was 10.3 (SD 1.8) with a range from 6 to 14. The clinical/practical component of the examination was marked out of 60 and the mean mark was 37.3 (SD 4.3) with a range from 26 to 46. In Fig. 2, all marks were converted to a percentage score for ease of comparison and graphical representation.

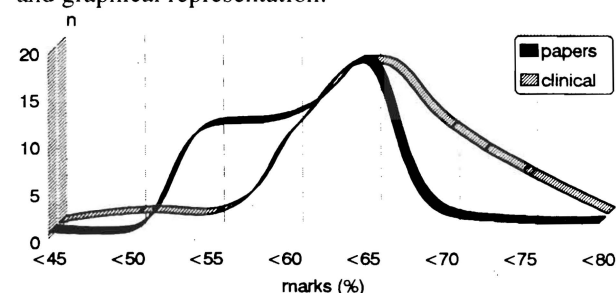


Fig. 2 - Frequency distribution of marks obtained in the written and clinical final qualifying examinations in medicine

The results obtained in the individual components of the examination were correlated with each other and with the total mark; this enabled insight into the

Table 1 - Analysis of the result of the final qualifying examination in medicine for 1995. Correlation between marks obtaining from the components of the examination themselves and the total mark.

Variable	Essay paper	MCQ paper	Clinical examination
Essay paper	1.000	-	-
MCQ paper	0.4631	1.000	-
Clinical exam.	0.3541	0.4757	1.000
Final mark	0.6456	0.7340	0.9082

relationship between the components themselves as well as their weighting and bearing on the final mark (Table 1).

The issue of reliability of examining methods was studied by calculating Cronbach's standardized alpha for all variables' and then on removing each variable separately from the equation. For purpose of analysis, the A-level result was also entered as a variable. Cronbach's alpha for all variables was 0.6625 (Table 2).

Table 2 - Analysis of the final examination results. Reliability of examining methods.

Variable removed	Effect on reliability	Cronbach's alpha
A-level results	increase	0.6944
paper 1 (essays)	decrease	0.6190
paper 2 (MCQ)	decrease	0.4746
clinical examination	decrease	0.5662

## Discussion

On the basis of their entry qualifications to the University of Malta Medical School, the group studied was not representative of the population at large and probably neither of the university student population. The medical course in Malta possibly attracts some of the most academically gifted of students undertaking tertiary education, in that only those candidates with passes at the highest grades in science subjects are accepted, according to explicit selection criteria. Demand for places always exceeds regulated supply, and competition for admission to the course of studies commencing every second year is tough; in 1990, 94 individuals applied for 53 places. All those accepted, offered very good A-level results as entry qualifications; the majority obtained grade A in three science subjects at one sitting, while most offered additional passes at A-level.

There was a statistically insignificant scant correlation between the entry qualifications and the composite mark obtained in the final examination in medicine, suggesting that entry qualifications are neither sufficiently predictive nor a reliable index of subsequent performance of medical students. Although the findings presented relate to one cohort in one university, similar results were obtained when multiple cohorts in several British universities were studied<sup>4</sup>. In both countries, entry to medical school is based on similar formal qualifications; and in the case of Malta this is exclusively so. With regard to other countries operating different systems for admission to medical school, similar findings hold. In Australia, most candidates are

in the top 2% of the matriculating population and also have studied science<sup>5</sup>. In the latter, and other countries, there is little link between achievement in science and/or overall academic achievement and later competence in clinical practice<sup>6,7</sup>.

Proficiency in humanities has been found to underpin successful performance at medical school<sup>8</sup>. The manner in which such proficiency is acquired should be subjected to deeper analysis. Subject spread and the inclusion of a background of humanities is important for effective medical practice, not least because this required good communication skills between doctor and patient and between health professionals themselves<sup>9</sup>. Indeed, the narrow interests of today's physicians are a source of concern in several countries<sup>10,11</sup>.

The methods used for the final examination in medicine were two 'pen and paper' tests and a clinical assessment. Of the three components, the essay paper was the least reliable determinant of the final result. On employing a mathematical model of reliability, and when results of the essay paper were removed, reliability only fell by 6.6%. When the MCQ results were removed from analysis, reliability fell by over 28%, suggesting that assessment through MCQ was by far the most reliable part of the examination. Removal of the results of the clinical examination from the model reduced reliability by 14.5%. These findings are not surprising given the nature of the tests as well as the methods used for evaluating and marking them.

An MCQ test usually samples a broad range of facts and students spend most of their time reading and thinking out the responses. With regard to essay questions only a small number of topics may be assessed and luck plays a large part in this type of examination. Essay questions require the student to mostly spend his time thinking and writing. While essay tests permit and occasionally encourage bluffing, MCQ tests permit and occasionally encourage guessing<sup>12</sup>.

Essay questions are easy to set while the construction of an MCQ paper is very time-consuming. Conversely, marking essay questions is time-consuming while MCQ responses are easily scored, possibly using an electronic scanner. More significantly, MCQ marks are perfectly reproducible while essay marks are often not. Indeed, the skill of the examiner marking students' answers often significantly determines the quality of the essay test, thus introducing an often unquantifiable variable. Criteria used by examiners to mark essays often vary in their consistency and it has been shown that correlations within and among examiners are low<sup>13</sup>. The very nature of the MCQ tests eliminates examiner bias and supposes only one variable; the candidates' responses.

The relative weight given to each of the three

components of the final examination on the award of the final mark was different. The results obtained in the clinical examination correlated best with the total mark. This is purposely and deservedly so, given that proficiency in basic clinical skills and person-to-person communication are essential attributes in those seeking the first formal medical qualification. The weakest correlation was that between the essay paper and clinical examination. The essay paper also exerted the least weight in the determination of the composite grade in medicine, and in this context it was reassuring to note that this component of the examination was also statistically the least reliable.

Essays do however have an important role in students' assessment, although perhaps not as part of the final examination. The discipline of formulating ideas coherently and sequentially is probably best tested through essays/projects assigned to students during the academic year. In such a case they would have full recourse to data retrieval facilities and would not be unduly constrained by time. The assessor would thereby gain insight into how the student marshalls facts critically, while the student would subsequently also gain through valuable feedback.

The final examination in medicine should have clearly defined objectives. It needs to be able to assess the ability to think critically about diagnosis and management, and to ensure that the candidate has a satisfactory base of factual knowledge. It also needs to assess the adequacy of basic clinical skills and the candidate's facility of communication.

A written-test format that requires comment on actual clinical situations, and one designed to test problem solving, offers the candidate the opportunity to demonstrate skills in pattern recognition and discrimination while also assessing the ability to think critically about patient management. Reliability of using this type of test however, supposes pre-determined model answers in order to increase consistency through decreasing examiner bias. With regard to this type of examination, candidates could be allowed to bring along to the examination hall whatever texts they may wish to consult. Although this may appear controversial, it has been shown that in some circumstances, it makes no difference if students are allowed to bring their notes to the examination<sup>14</sup>. It certainly helps to allay their anxiety, and indeed in real-life situations, doctors very often consult books and data-bases in order to improve the quality of the care they give. Both breadth and depth of knowledge are efficiently tested by means of the MCQ test and this format also ensures reliable, impartial and reproducible scoring. In addition, it facilitates statistical analysis, not only of individual results, but also of the examination itself and indirectly of the effectiveness of teaching.

Written tests do not, however, measure clinical competence and the task of measuring clinical performance is probably the most difficult and taxing of the whole examination. It is also the most important part of the examination, not least because the personal contact also affords assessment of attitude and of ability to communicate. As shown in this paper, and justifiably so, this part of the examination is given most weight at the Malta medical school. The clinical examination assesses perception of essential clues, their appreciation

and the significance attached to them. It enables the candidate to present an argument logically, backed by observations and facts. A significant defect of this type of examination lies in observer bias, and it has been shown that reliability within and among examiners is not always high<sup>15</sup>. As six examiners assess the candidate in the final three-part clinical medicine examination in Malta, bias tends to be diminished. Notwithstanding this, there is still scope for more objective scoring and marking systems. Indeed, it is with regard to all parts of the examination that assurance of quality must be of primary concern; this also supposes openness to scrutiny and the potential that examiners are accountable for their decisions.

In summary, the findings of this paper suggest that overall there is room for improving assessment methods regarding the qualifying examination in medicine. A review of methods and procedures, with the dual purpose of decreasing bias and increasing specificity and sensitivity of the examination will not only be of benefit to candidates but ultimately also to the University of Malta.

### Acknowledgements

*We thank Prof. A. Buhagiar for help with statistics, and Prof. FF Fenech and Mr. J. Bartolo for facilitating this study. Before submitting this paper we had the benefit of comments from two persons who have to remain anonymous as we cannot compromise them with our opinions; to both these gentlemen we are grateful.*

### References

1. Scriven M. The methodology of evaluation. in Stake RE, (ed). Curriculum evaluation. American Education Research Association. monograph series on evaluation No. 1 Rand McNally, Chicago, Illinois. 1967.
2. Malta Medical Students Association. A new approach to setting standards; course evaluation by students. Saliba K (ed). Unpublished data on file. 1994.
3. Malta Medical Students Association. Medical education evaluation report for the intermediate clinical year 1994 and clinical year 1994-1995. Galea I (ed). Unpublished data on file. October 1995.
4. Mahwinney BS. The value of ordinary and advanced level British school leaving examination results in predicting medical students' academic performance. *Med Educ* 1976; 10: 87-89.
5. Neame RLB, Powis DA, Bristow T. Should medical students be selected from recent school-leavers who studied science? *Med Educ* 1992; 26: 433-40.
6. Dickman RL, Samacki RE, Schimpfhauser FT, Katz LA. Medical students from natural science and non-science undergraduate backgrounds: similar academic performance and residency selection. *JAMA* 1980; 250:6-09.
7. Woodward CA, McAuley RG. Can the academic background of medical graduates be detected during internship? *Can Med Assoc J* 1983; 129: 567-69.
8. Lipton A, Huxham GJ, Hamilton D. Predictors of success in a cohort of medical students. *Med Educ* 1984; 18: 203-10.
9. Rolfe I, Pearson S, Powis DA, Smith AJ. Time for review of admission to medical school? *Lancet* 1995; 346: 1329-33.
10. Nash IS, Pasternak RC. Physician educate thyself. *JAMA* 1995; 273: 61-2.

11. Wiedersheim R, Muller-Busch HC. A new approach to medical education. In Kaplun A, (ed). Health promotion and chronic illness. Copenhagen, WHO Regional Office for Europe 1992; 351-53.
12. Ebel RL. Essentials of educational measurement. Prentice-Hall, New Jersey 1972.
13. Bull GM. An examination of the final examination in medicine. Lancet 1956; 2: 368-72.
14. Trigwell K. The crib card examination system. Assessment Evaluation Higher Educ 1987; 12: 56-65.
15. Levine HG, McGuire CH. The validity and reliability of oral examinations in assessing cognitive skills in medicine. J Educ Measurement 1970; 7: 63-74.

The copyright of this article belongs to the Editorial Board of the Malta Medical Journal. The Malta Medical Journal's rights in respect of this work are as defined by the Copyright Act (Chapter 415) of the Laws of Malta or as modified by any successive legislation.

Users may access this full-text article and can make use of the information contained in accordance with the Copyright Act provided that the author must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the prior permission of the copyright holder.

This article has been reproduced with the authorization of the editor of the Malta Medical Journal (Ref. No 000001)