CONSISTENCY OF UNIVERSITY EXAMINATIONS

AN ANALYSIS OF SOME EDINBURGH MEDICAL RESULTS

BY

W. GILBERT MILLAR, M.B., Ch.B.

(From the Department of Pathology, University of Edinburgh)

Considerable disquiet has been expressed of late as to the reliability of examinations as a guide to the ability of the candidate and his fitness for further study. The work of Hartog and Rhodes,¹ and to a lesser degree the results published by the Scottish Council for Research in Education² (both parties working on behalf of the International Examination Inquiry) certainly provide a sound basis for the doubts now so freely expressed.

On the other hand, the general impression of most of the more experienced examiners in the medical subjects in the University of Edinburgh (and no doubt elsewhere) is that, broadly, the examinations do give a pretty fair estimate of a student's capabilities, and that anomalies such as are disclosed by Hartog and Rhodes do not occur, or, if they do, only to such an extent as to be negligible.

While a full analysis on the lines of that carried out on behalf of the English Committee of the International Inquiry was not practicable, an investigation has been made into the correlation existing in a number of marks recorded in certain medical subjects.

Material Analysed

The material which forms the basis of this analysis consists of two series of marks.

(A) The marks obtained by candidates taking the six subjects which constitute the combined third-year class examination in the course for M.B., Ch.B. at Edinburgh University, for the six years 1929-30 to 1934-5 inclusive.

(B) The marks obtained in the various professional examinations by those candidates who took the combined examination for the first time in the two years 1931-2 and 1932-3. These two groups were chosen as including those candidates whose whole course would normally extend over the period covered by the first part of the analysis.

Since the third-year teaching and examining in Edinburgh University is arranged on a plan rather different from that obtaining in most other schools, it seems desirable, before considering the figures, to deal briefly with the methods employed.

Before the year 1928-9 the various third-year subjects were taught independently, and separate class examinations were held, on the results of which the candidate's right to sit for the third professional examination was judged. In 1928-9 an important modification was introduced, largely owing to the influence of the late Lorrain Smith. From that year onwards the systematic teaching in pathology, bacteriology, pharmacology, therapeutics, medicine, and surgery has been co-ordinated so that the student is, so far as possible, taught all aspects of a particular disease at the same time. For example, if in medicine the subject is valvular disease of the heart, then in pathology endocarditis will be dealt with, in bacteriology streptococci and the bacteriology of acute rheumatism, and in pharmacology and therapeutics the actions and uses of the digitalis group. The surgical element in cardiac disease being relatively unimportant, the surgeons would utilize this time to cover some purely surgical subject.

The examinations in connexion with this co-ordinated teaching are also co-ordinated and are held in spring and summer. At each period two papers are set, one of three hours and one of two hours. In each case the departments concerned are grouped together variously and share questions. For example, a question may be set ask-

ing for the pathology of pyelonephritis, the surgical treatment of renal calculus, and the bacteriology of urinary infections, while medicine and pharmacology and therapeutics might share a question on cardiac disease. Each part of the question is written in a separate book and marked by the department concerned.

The administration of the joint examination is carried out by each department in turn. Before each examination the heads of the six departments and the lecturer managing the examination for the current year meet to set the papers, and after each examination meet again to consider results. The marks attained by each candidate in each subject are set out in full on a schedule, and his claims to a "D.P." (certificate of having duly performed the work of the class which admits him to the professional examination) are considered jointly by the six examiners. In most cases, of course, there is little doubt one way or the other, but the borderline cases are fully considered. The general impression gained by all who have had to do with the joint examination is that a candidate's "form" from one subject to another is remarkably constant, much more so, indeed, than might be expected considering the diverse character of the subjects studied.

To confirm or correct the general impression of uniformity of performance, correlation coefficients have been calculated between each of the six subjects for each of the six years under review. The year of starting the scheme (1928-9) has been omitted for two main reasons. First, the scheme was in a tentative form, and secondly the records kept were not conveniently accessible. By taking six years the marks considered include one turn of administration for each department. Since in any one year there are a certain number of hitherto unsuccessful candidates repeating the course, and since the marks gained by them in their second or subsequent attempts are not comparable with those secured at a first try, only first attempt marks are included in the material. Only the marks of those who attended the whole of each examination have been taken.

The second part of the analysis is an attempt to assess the prognostic value of the combined examination, and to compare it with that of the preceding and succeeding professional examinations. From the students sitting the combined examination in the years 1931-2 and 1932-3, 231 were taken who had attended their entire medical course at Edinburgh University, and for whom, therefore, the marks were readily available. The figures taken were in all cases those of the first attempt at each examination, the marks in subsequent attempts being disregarded for the reason already considered. It is scarcely necessary to point out that the 231 students all sooner or later obtained pass marks in the earlier examinations, and the figures are therefore to some slight extent selected, since the very weakest or least persistent students tend to drop out and do not reach the final at all.

The figures correlated consist in the average mark secured in the respective examination taken as a whole. In the first professional examination are averaged written and oral (and/or practical) in each of botany, physics, chemistry, and zoology; in the second, written and oral (and/or practical) in physiology and anatomy; in the combined examination the average of written papers in pathology, bacteriology, pharmacology, therapeutics, medicine, and surgery; in the third professional, written and oral (and/or practical) in pathology (including bacteriology) and pharmacology (including therapeutics); in the final, written and oral in forensic medicine and public health (taken at the end of the fourth year but included in the final), midwifery (including gynaecology), medicine and surgery, together with clinical midwifery (including clinical gynaecology), clinical medicine, and clinical surgery.

Method

The correlation coefficients were calculated in the usual manner, but Sheppard's correction was not applied. In many cases there was high contact at one end only of the distribution and in a few at neither. For the

sake of uniformity it was thought better not to apply the correction in any of the cases. The actual distribution of the marks for the forming of the correlation tables was very greatly facilitated by the use of Paramount punched cards on which the various marks were entered at the outset. In the case of the correlations between the subjects of the combined examination, where the range of marks was fairly wide and the frequencies moderate, the figures were arranged in ten-mark groups. In the case of the professional and average combined marks in the second series, the averaging had evened out the irregularities and somewhat diminished the range. Moreover, a larger number of pairs were available, so that a five-mark grouping was found to be preferable.

Results

The results of the analysis are given in the following tables. Tables I to VI show the values of the correlation coefficient r between the six subjects of the combined third-year examination for the six years 1929-30 to 1934-5 respectively, along with the total frequency, N , for each year.

TABLE I.—1929-30. $N = 185$

	Pathology	Bacteriology	Pharmacology	Therapeutics	Medicine	Surgery
Pathology ...	—	0.64±0.03	0.43±0.04	0.37±0.04	0.48±0.04	0.53±0.04
Bacteriology ...	0.64±0.03	—	0.52±0.04	0.59±0.03	0.50±0.04	0.59±0.03
Pharmacology ...	0.43±0.04	0.52±0.04	—	0.60±0.03	0.50±0.04	0.42±0.04
Therapeutics ...	0.37±0.04	0.59±0.03	0.60±0.03	—	0.39±0.04	0.49±0.04
Medicine ...	0.48±0.04	0.50±0.04	0.50±0.04	0.39±0.04	—	0.49±0.04
Surgery ...	0.53±0.04	0.59±0.03	0.42±0.04	0.49±0.04	0.49±0.04	—

TABLE II.—1930-1. $N = 157$

	Pathology	Bacteriology	Pharmacology	Therapeutics	Medicine	Surgery
Pathology ...	—	0.45±0.04	0.48±0.04	0.38±0.05	0.57±0.04	0.45±0.04
Bacteriology ...	0.45±0.04	—	0.56±0.04	0.50±0.04	0.56±0.04	0.53±0.04
Pharmacology ...	0.48±0.04	0.56±0.04	—	0.63±0.03	0.61±0.03	0.54±0.04
Therapeutics ...	0.38±0.05	0.50±0.04	0.63±0.03	—	0.55±0.04	0.51±0.04
Medicine ...	0.57±0.04	0.56±0.04	0.61±0.03	0.55±0.04	—	0.50±0.04
Surgery ...	0.45±0.04	0.53±0.04	0.54±0.04	0.51±0.04	0.50±0.04	—

TABLE III.—1931-2. $N = 139$

	Pathology	Bacteriology	Pharmacology	Therapeutics	Medicine	Surgery
Pathology ...	—	0.42±0.05	0.57±0.04	0.60±0.04	0.51±0.04	0.38±0.05
Bacteriology ...	0.42±0.05	—	0.46±0.05	0.43±0.05	0.52±0.04	0.46±0.05
Pharmacology ...	0.57±0.04	0.46±0.05	—	0.55±0.04	0.53±0.04	0.53±0.04
Therapeutics ...	0.60±0.04	0.43±0.05	0.55±0.04	—	0.59±0.04	0.46±0.05
Medicine ...	0.51±0.04	0.52±0.04	0.53±0.04	0.59±0.04	—	0.58±0.04
Surgery ...	0.38±0.05	0.46±0.05	0.53±0.04	0.46±0.05	0.58±0.04	—

TABLE IV.—1932-3. $N = 184$

	Pathology	Bacteriology	Pharmacology	Therapeutics	Medicine	Surgery
Pathology ...	—	0.45±0.04	0.43±0.04	0.40±0.04	0.47±0.04	0.34±0.04
Bacteriology ...	0.45±0.04	—	0.52±0.04	0.34±0.04	0.45±0.04	0.32±0.04
Pharmacology ...	0.43±0.04	0.52±0.04	—	0.54±0.03	0.60±0.03	0.49±0.04
Therapeutics ...	0.40±0.04	0.34±0.04	0.54±0.03	—	0.51±0.04	0.42±0.04
Medicine ...	0.47±0.04	0.45±0.04	0.60±0.03	0.51±0.04	—	0.54±0.03
Surgery ...	0.34±0.04	0.32±0.04	0.49±0.04	0.42±0.04	0.54±0.03	—

TABLE V.—1933-4. $N = 186$

	Pathology	Bacteriology	Pharmacology	Therapeutics	Medicine	Surgery
Pathology ...	—	0.46±0.04	0.52±0.04	0.53±0.04	0.55±0.03	0.62±0.03
Bacteriology ...	0.46±0.04	—	0.46±0.04	0.40±0.04	0.45±0.04	0.48±0.04
Pharmacology ...	0.52±0.04	0.46±0.04	—	0.46±0.04	0.55±0.03	0.51±0.04
Therapeutics ...	0.53±0.04	0.40±0.04	0.46±0.04	—	0.55±0.03	0.55±0.03
Medicine ...	0.55±0.03	0.45±0.04	0.55±0.03	0.55±0.03	—	0.60±0.03
Surgery ...	0.62±0.03	0.48±0.04	0.51±0.04	0.55±0.03	0.60±0.03	—

TABLE VI.—1934-5. $N = 197$

	Pathology	Bacteriology	Pharmacology	Therapeutics	Medicine	Surgery
Pathology ...	—	0.32±0.04	0.36±0.04	0.52±0.03	0.43±0.04	0.25±0.04
Bacteriology ...	0.32±0.04	—	0.40±0.04	0.48±0.04	0.40±0.04	0.18±0.05
Pharmacology ...	0.36±0.04	0.40±0.04	—	0.45±0.04	0.46±0.04	0.27±0.04
Therapeutics ...	0.52±0.03	0.48±0.04	0.45±0.04	—	0.50±0.04	0.23±0.05
Medicine ...	0.43±0.04	0.40±0.04	0.46±0.04	0.50±0.04	—	0.43±0.04
Surgery ...	0.25±0.04	0.18±0.05	0.27±0.04	0.23±0.05	0.43±0.04	—

It will be seen that the values for r range round about + 0.5 for the first five years, dropping slightly for the sixth year to about + 0.4. Since the probable error attached to this last value is only about 0.04 it will be seen that a very significant correlation of high degree attaches to these figures. It should be noted that the probable errors are quoted for the coefficients not because they form the best guide to significance in the case of correlations but because they provide the readiest method of expressing it. The significance of the results has been checked both by the use of "t" and of "z," and the results do not differ materially from those obtained from the probable error. This is to be expected in view of the values of r obtained (not very close to 1.0) and of the fairly high total frequency.

The high values of r obtained indicate that the performance of the candidate, although judged in six different subjects, and in the case of most of these independently by three or four different examiners in each subject, remains very constant. It is straining probability too much to suggest that this constancy is due to all the examiners forming the same misjudgement of the candidate's abilities. The only alternative is that this examination does give a consistent estimate of the candidate's powers.

The remaining results illustrate the prognostic value of the examination. Table VII shows the correlation coefficients derived from the marks of the first, second, third, and final professional examinations, and the combined examination.

TABLE VII.—1931-3. $N = 231$

	Final	Combined
Final ...	—	0.66±0.02
Combined ...	0.66±0.02	—
First Professional ...	0.54±0.03	0.51±0.03
Second " ...	0.68±0.02	0.65±0.02
Third " ...	0.64±0.02	0.69±0.02

Excluding those for the first professional, the values for r range round about + 0.66 with a probable error in the neighbourhood of 0.02. The correlation is very high and very significant. The comparatively low values for the first professional (which are also really very high) are no doubt in part to be explained by the fact that at this early period the candidate has had relatively little practice in writing examinations, as well as that the non-medical character of the subjects not infrequently fails to arouse the serious attention of the student.

To those acquainted with statistics the results speak for themselves and need no discussion, but a brief consideration of one or two of the correlation tables yields some points of interest which should be equally plain to those with no great trust in statistical methods. Table VIII shows the arrays of marks secured in the final and in the combined examination.

TABLE VIII.—Final

Percentages	45-50	50-55	55-60	60-65	65-70	70-75	75-80	Total
Combined	30-35	1						1
	35-40	4	4	4				12
	40-45	2	9	12	1			24
	45-50		13	14	10			37
	50-55		9	33	28			70
	55-60		2	17	24	4	2	49
	60-65			3	14	6	2	26
	65-70		1	1	1	5	2	10
	70-75					2		2
Total	7	38	84	78	15	8	1	231

Taking the lower marks first it will be seen that of the thirteen candidates securing less than 40 per cent. in the combined, none reached an average of 60 per cent. in the final and only four secured from 55 per cent. to 59 per cent. Of the seven who obtained less than 50 per cent. in the final none had secured 45 per cent. in the combined and only two had reached 40 per cent. When it is pointed out that the pass mark is 40 per cent. in the combined and 50 per cent. in the final the agreement is even stronger. It must be remembered that in addition to the candidates here considered a certain number below 40 per cent. in the third year became discouraged and gave up. The agreement is even more striking at the upper end. Of the twenty-four candidates who attained the respectable figure of 65 per cent. and over in the final, none had secured less than 55 per cent. in the combined and only six had obtained less than 60 per cent. Moreover, out of the thirty-eight who reached 60 per cent. or over in the combined—that is, the group of really good students from the third-year point of view—only five obtained less than 60 per cent. in the final. Fifteen secured between 60 per cent. and 64 per cent., while eighteen secured 65 per cent. or more. It can thus be claimed that the result of the combined examination is a very fair guide to the candidate's chances in the final, be he strong or weak.

Table IX shows the arrays for the combined and for the third professional examination.

TABLE IX.—Combined

Percentages	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	Total
Third Professional	35-40			1						1
	40-45		4	2	2					8
	45-50	1	4	10	12	9	1			37
	50-55		8	8	12	25	5	1		60
	55-60			2	9	22	21	7	1	62
	60-65				1	11	19	15	5	51
	65-70					3	3	2		8
	70-75							1	1	2
	75-80							1	1	2
Total	1	12	24	37	70	49	26	10	2	231

The correspondence is high here also. Of the thirteen candidates who failed to secure 40 per cent. in the combined, only eight reached 50 per cent. in the professional and none reached 55 per cent. It must be remembered that some of these eight would have sat the combined

a second time and so qualified to sit the professional. At the other end of the scale, of the twelve candidates who reached or exceeded 65 per cent. in the professional none had secured less than 55 per cent. in the combined and only three had failed to reach 60 per cent. Of the twelve candidates who secured 65 per cent. or more in the combined only one secured less than 60 per cent. in the professional and only six less than 65 per cent.

Considerations of space preclude a detailed analysis of all the correlation tables, but they show the same features in greater or smaller degree, as of course is implied in the high values of r .

Summary

1. Correlations have been calculated between the six subjects of the combined examination for each of six years—that is, for a total of 1,048 candidates. The correlation coefficient, r , ranges round about + 0.5 with a probable error of the order of 0.04, indicating a high degree of correspondence in the results.

2. Correlations have also been calculated between the final and other professional examinations, and the combined examination for a group of 231 students. The coefficients obtained range round about + 0.6, with probable error in the neighbourhood of 0.02, showing a very high degree of correspondence.

3. Study of the correlation tables reveals a high degree of correspondence between the results of the combined examination and of the third and final professional examinations at both the upper and the lower ends of the scale of marks.

4. From the results obtained the conclusion is justified that the examinations considered form a reliable guide to the students' academic capabilities.

My acknowledgements are due to the Dean and Faculty of Medicine of Edinburgh University for access to the marks forming the material for this investigation.

REFERENCES

- Hartog, P., and Rhodes, E. C.: *An Examination of Examinations*, second edition, 1936.
- Scottish Council for Research in Education: *The Prognostic Value of University Entrance Examinations in Scotland*, 1936.

L. Hill (*Journ. of Physiol.*, July, 1936, p. 45P) discusses the effect on the lungs of breathing through a narrow orifice. The difficulty of breathing through the narrow orifice of a short tube is increased by exposing the face to heat rays and is eased by the cooling effect of a fan. These effects are not psychological, but are more likely due to reflex narrowing of the air passages by heat and their opening by cooling. When the orifice of the tube is so small that a breaking-point in respiration soon occurs, the carbon dioxide content of the alveolar air is not notably increased, while the oxygen percentage is reduced to 9-10. When inhaling oxygen-enriched air through the breathing-tube, the alveolar oxygen percentage is raised to a high level, but the breaking-point is little extended, and the greater difficulty of breathing on exposure to heat rays and the easement on cooling the face remain. Breathing oxygen-enriched air before and not during the experiment prolongs the breaking-point. Making expiration free and obstructing only inspiration greatly reduces the difficulty of breathing. In explanation it is suggested that breathing through a narrow orifice greatly impedes the pulmonary circulation, and that owing to this and the consequent oxygen lack the respiratory centre fails. Breathing oxygen during the experiment helps little because the pulmonary circulation is impeded; breathing oxygen beforehand helps because the blood and the respiratory centre are well supplied with oxygen in advance. When a free inspiration is taken before sampling the alveolar air, the alveolar air contains much less oxygen than normal, because the obstructed blood, when released to pass freely, takes up more oxygen than normally. Impedance of the pulmonary circulation must be of clinical importance in cases of obstructed breathing. The capacity to breathe through orifices of varying sizes can be used to assess fitness.