J. Natn. Sci. Coun. Sri Lanka 1975 3(1) : 31-39

On Examinations

T. D. M. A. SAMUEL
Department of Engineering Mathematics, University of Sri Lanka, Peradeniya Campus, Peradeniya, Sri Lanka.

(*Paper accepted :* 17 *March* 1975)

Abstract : The parameters possibly influencing the performance at examinations are discussed. After a simple analysis, the results of a particular examination are discussed in terms of these parameters. This discussion reveals the need for the definition of a 'Standard Performance Range'.

1. Introduction

The main purpose of this paper is to increase the understanding of the machinery of examinations. In order to achieve this, the parameters possibly influencing the 'performance' at an examination are postulated. This is followed by a general discussion of these parameters and a discussion of general remedial measures against unfavourable influences of these parameters at examinations. A need for the introduction of definitions of 'performance', etc. is immediately felt and these are provided in the Appendix.

This is followed by a very simple analysis of the Final Part I Examination in Engineering of the University of Sri Lanka, Peradeniya Campus held immediately after a change in Regulations and Syllabus. The analysis reveals certain inconsistencies. Guide lines are therefore suggested for the maintenance of a certain degree of consistency.

2. Parameters influencing end of year Examination

2.1.

In this paper, examination is viewed in its role in ranking students belonging to a particular batch and in deciding on pass/fail. Feed back information for teaching, etc. are not considered. The rank and pass/fail of candidates depend on their performance in individual subjects and also on the examination as a whole. The performance in a particular subject at an examination depends largely on :—

(i) standard of paper
(ii) standard of marking
(iii) attitude of examiner towards candidates
(iv) examination temperament of candidates
(v) attitude of student towards subject
(vi) nature of the subject
(vii) foundation in allied subjects in the previous years.

The first three are in the domain of influence of the examiner, the fourth and fifth in that of the candidate and the last two depend on the subjects. At this stage, it does not appear to be possible to determine the exact degree to which each of the above factors influence performance. Factors like (vi) which are not totally independent of the other factors would make this kind of analysis extremely difficult if not impossible. But an analysis of the performance at an examination, influenced to varying degrees by these factors, can increase our understanding to enable us to formulate in the future a reasonable Mathematical model.

In this context, if it is assumed that (i) – (v) and (vii) are conducive to a 'standard performance', any unusual performance (see Appendix) in any particular subject or subjects *may be attributed* to the difference in the nature of the subject or subjects. The existence of this kind of difference is seen from the example of distribution curves for subjects in a School Certificate Examination obtained by Crofts and Jones[1] as shown in Figure 1.

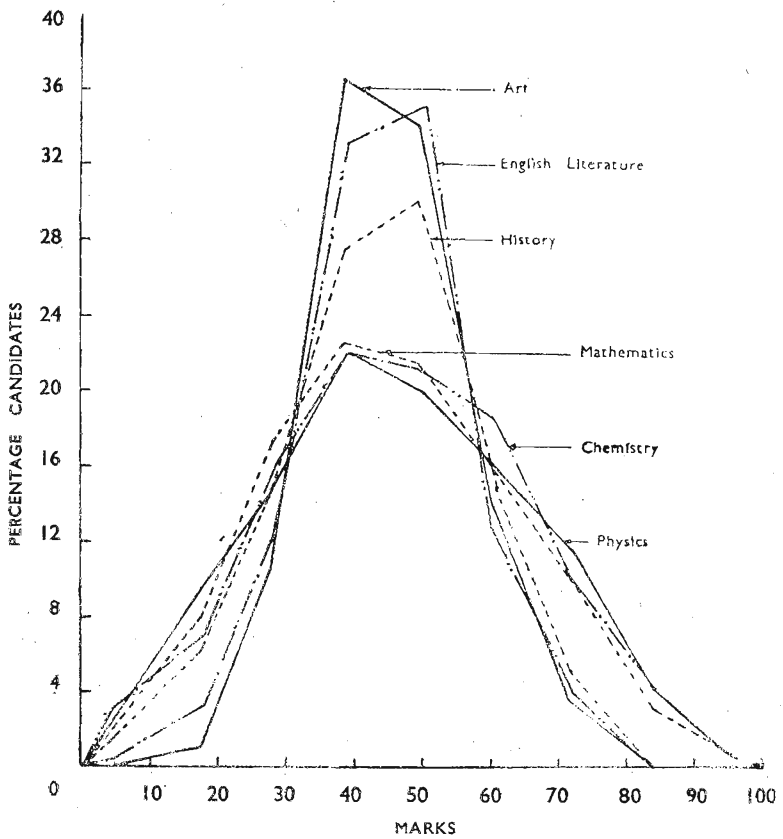

Figure 1. Shows two distinct groups of subjects.

In this example, the subjects fall into two distinct groups. The curves are flatter for subjects like Mathematics and Science and they are peaked towards the centre for subjects like History and Art. In the latter case a very large number of candidates get marks in a range about 45. In the former, there is a wider spread with the possibility of very high and very low marks. These differences in curves are reported by Crofts and Jones to be persistent. Furthermore, it is found that the coefficient of rank correlation (see Appendix) between examination and class performance (continual assessment) is as small as 0.75 compared with 0.85 for the former (flatter curves). These are due to inherent differences in the subjects.

In the event of the existence of two distinct subject groups as described above, the following difficulties arise :—

(i) Assume that in an examination only one subject—subject A (say)—belongs to that group where it is difficult to score high marks. If this examination is a competitive one, in which case the rank of candidates is important, then those not offering A will be at a definite advantage over the others.

(ii) This can also affect the 'doubtful cases' because subject A with only low marks will reduce the average which has an influence on the decision of the board of examiners.

A conscious effort to remove these difficulties must be evident in any examination system. Having established the nature of the subjects in an examination we can lay guide lines for consistency in (i)—(v).

2.2. Examiner

(i) and (ii) are dependent on the examiners. Assuming that (iii)—(vii) do not have any adverse influence on the examination, it is reasonable to expect a fair degree of consistency in performance in a particular subject. This would be true provided the different batches are statistically equivalent and that the standard of the subject has not changed drastically. If this position is granted, a useful guide would be to compare the performance in any examination against a 'model performance range'. This is discussed in the sections to follow.

2.3. Candidates

Poor performance due to (iv) can be alleviated by continual assessment of the students in the different subjects. It can take the form of tests in course units or its equivalent in the form of tutorials. On the results of this, and the end of year examination, it is possible to determine the coefficient of rank correlation between continual assessment and the final examination. If in a particular examination this 'coefficient' is large enough, then a student who has just failed to obtain the pass mark can be pushed up on the strength of the continual assessment marks.

It is difficult at this stage to formulate guide lines against failings in (iii) and (v), but it is hoped that the elimination of discrepencies in (i), (ii), (iv), (vi) and (vii) will alleviate failings in (iii) and (v).

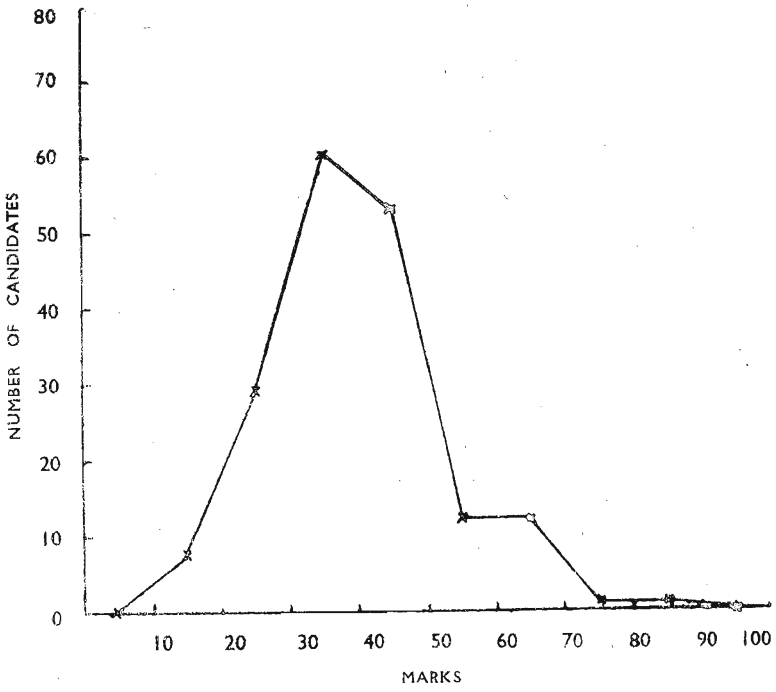

Figure 2. Frequency distribution for July 1969 Examination.

3. Analysis of Part I (New Regulations) results

Figure 2 gives the frequency distribution curve for the averages obtained by the students in the July 1969 examination. This was the first time that this examination was held. The parameters for the averages are as follows :

Arithmetic Mean = 39.0	UQ-Median = 7
Median = 38.0	Interquartile Range = 14
Mode = 34.5	Quartile Deviation = 7
Lower Quartile (LQ) = 31	Standard Deviation = 12.45
Upper Quartile (UQ) = 45	Third Moment = 0.99*
Median-LQ = 7	Fourth Moment = 3.69*

*Third and Fourth Moments for a Normal Distribution are 0 and 3 respectively.

These show that the average representative of the group is 39% and the average most frequently awarded is 34.5%. But the distribution is fairly symmetric about 38%. Thus a symmetric distribution has been obtained at the cost of an unusually

very poor performance. The performance at this examination as measured by the averages is shown by curve A of Figure 3. The performance in the August 1970 examination, of only those who failed the July 1969 examination, is shown by curve B. It is seen that the performance is better in the whole range of the averages. This is very unusual assuming consistency in (i)—(vii). A slightly better performance in a small range about 40%, would have been a reasonable result. This is seen in the case of the August 1970 first time failures repeating the examination in March 1971 as shown by curves C and D of Figure 3.

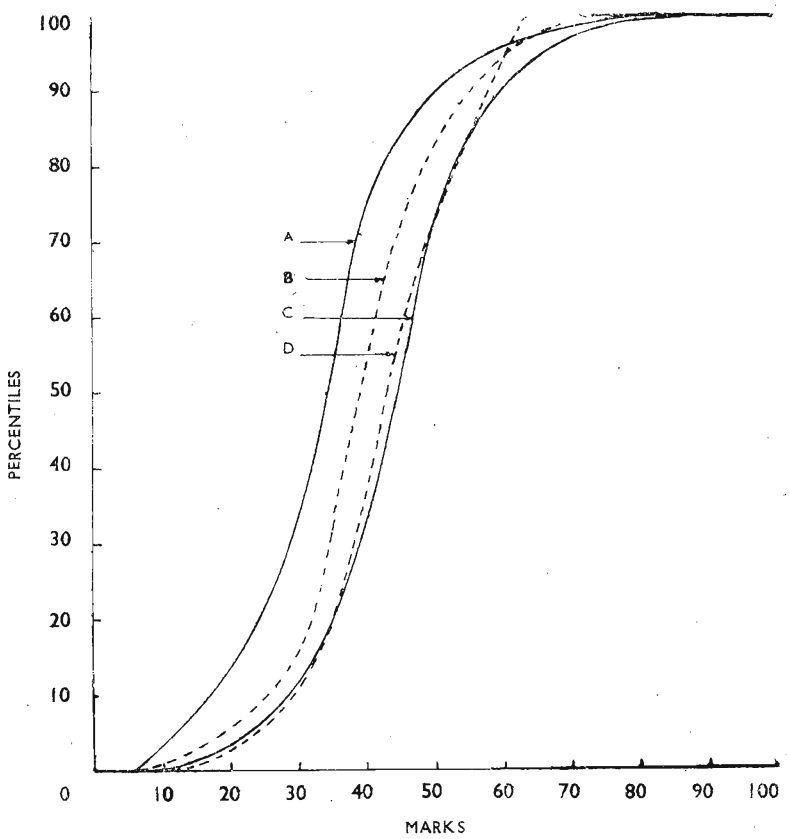

Figure 3. A—Part I, July 1969 ; B—Failures of Part I July 1969 sitting examination in August 1970 ; C—Part I, August 1970 first attempt only ; D—Part I Repeat Examination March 1971.

The performance at the August 1970 examination is a big improvement on the July 1969 examination. The performance in each of the five subjects at the July 1969 examination is shown in Figure 4. All except subject 4 have comparable performance. The general performance of the batch determined by the percentile

curve of the averages AV as shown in the figure has been greatly influenced by subject 4. Figure 5 shows the performance curves in the different subjects of the August 1970 examination. The performance range of the subjects is fairly wide and there is no substantial agreement in performance in any two subjects. There are extreme qualities in performance like subject 4 at one end and subjects 1 and 5 at the other. With this wide variety in quality of performance, it is very unlikely that tests in all subjects have been successful in the just fulfilment of their roles. This shows the need to define a 'Standard Performance Range'.

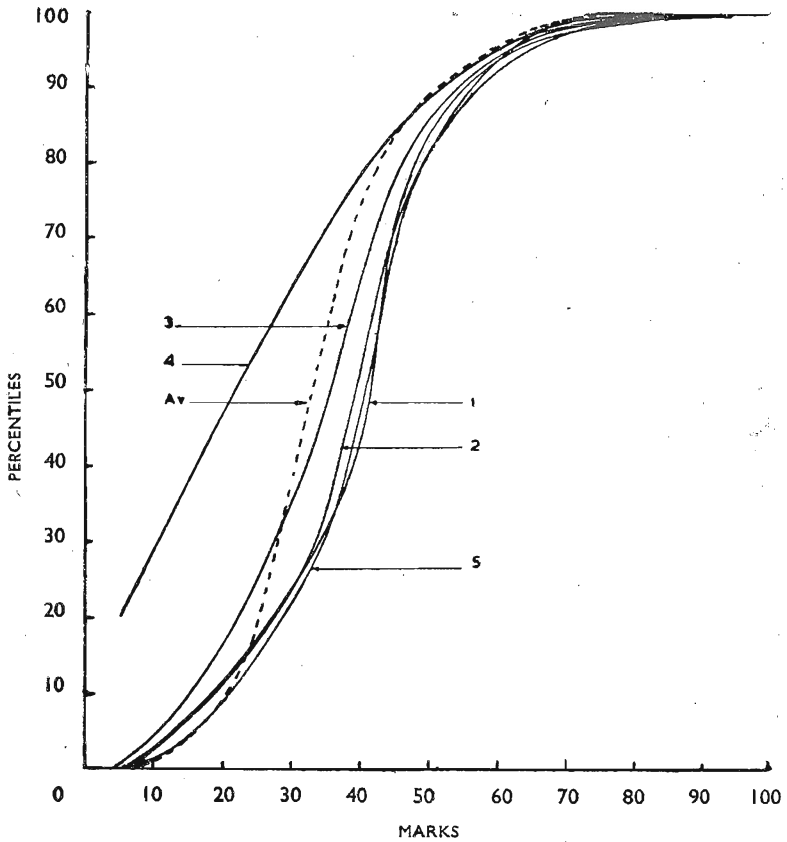

Figure 4. The percentile curves for subjects 1 to 5 and average AV of the July 1969 Examination.

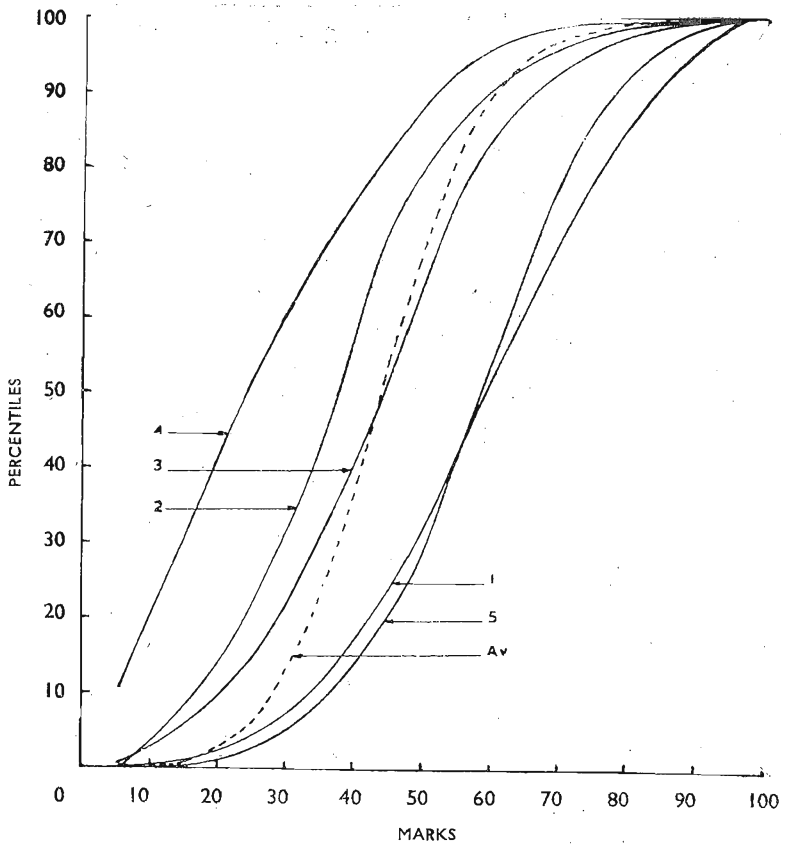

Figure 5. The percentile curves for subjects 1 to 5 and average AV of the August 1970 Examination.

4. Suggestions

A 'Standard Performance Range' for the Part I New Regulations examination as a whole may be obtained from the analysis of the averages of an 'equivalent examination' the results of which over a long period of time, are already available. When any examination performance falls outside this range i.e. when the performance is unusual, it is advisable to analyse the results in order to ascertain which of the parameters has influenced it. If it is possible to establish this, then a method of bringing the performance to within the 'Standard Performance Range' — *standardisation* — may be found. This may also be achieved by determining the 'Standard Performance Range' for each subject ; and ensuring — if possible, after analysis of results — that the performance in each subject falls within its subject performance range. This in itself will not ensure that the performance in the whole examination will fall within

the Performance Range of the whole examination, if one of the subjects is influenced by parameter (vi). In this case, in order to surmount difficulties (i) and (ii) discussed in section 2, the marks in this subject should be weighted. The weightage can be determined using the percentile curve of this subject and a percentile curve—to be chosen— in the 'Standard Performance Range'. If it is found that it is not possible to do any of these in order to correct an unusual examination performance and therefore since the examination has failed in its role, then the only course open is to repeat the examination. This defines a *repeat examination.*

5. Conclusions

The traditional end of year examination is still the only method of assessment in many institutions in Sri Lanka. Even if other methods of assessment are introduced, there are compelling reasons for the continued existance of the traditional examination as a major assessment method. It is therefore essential that it is made more reliable. The gradual evolution of a 'Standard Performance Range' is therefore of paramount importance to any educational institution. The author expects the 'Standard Performance Range' to be different for different institutions.

A major assumption in this paper is that the different batches of students are statistically equivalent. In this context, the 'Standard Performance Range' can indicate any sudden change in quality of students entering the institution. The author hopes to show a method of determining a 'Standard Performance Range' in a subsequent article.

Appendix

Definitions

The mth *percentile*, Pm, of a frequency distribution is the value of the variable x, that corresponds to the cumulative frequency, m per cent of N.

In the problem considered, x = percentage marks/average, m = percentage of students obtaining a certain percentage of marks/average, N = total number of students.

Ranked Data

Data arranged in order of relative magnitude or importance is called Ranked Data.

Coefficient of rank correlation

This is defined by the formula (Spearman formula)

$$r_R = 1 - \frac{6 \Sigma d^2}{N(N^2 - 1)}$$

where d is the difference between any paired ranks. The value is restricted to the range $-1 \leq r_R \leq +1$. No assumption concerning underlying distribution of variables is made.

New Definitions

The *performance* of a group of students in any subject or in the whole examination is defined in terms of the Percentile Curve (Orgive Curve). Two such curves can define the boundaries for a range of performance. Performance in a subject/examination is said to be *Standard* if the Percentile Curve falls within this range. If it falls outside this range then performance is said to be *Unusual.*

Acknowledgements

The author is grateful to Mr. V. Sanmuganayagam of the Faculty of Engineering, University of Sri Lanka, Peradeniya Campus for spending his valuable time in drawing the figures.

References

1. CROFTS, J. M. & JONES, D. C. (1927) *Secondary school examination statistics*, p. 49, London : Longmans, Green & Co. Ltd.