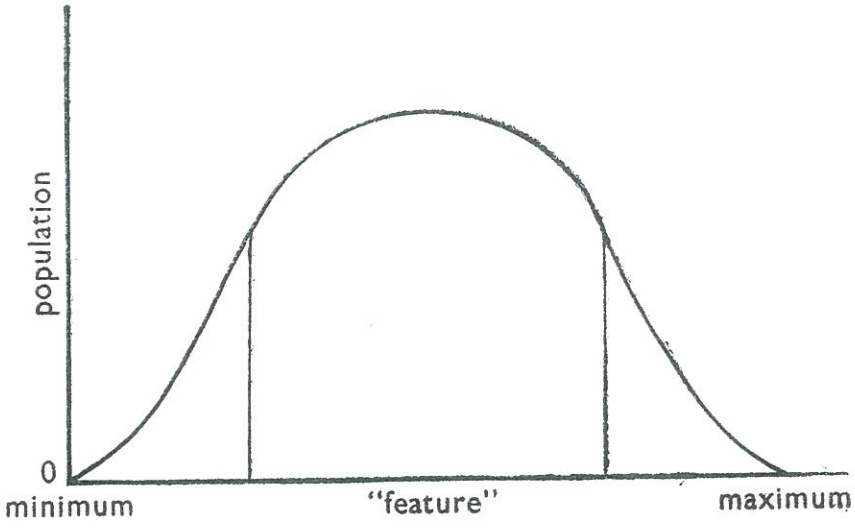
George (1961d) Bulletin of the Central Institute of English (Hyderabad) (1961) Issue 1

STATISTICS, SCHOOL-LEAVING EXAMINATIONS AND LANGUAGE TEACHING

H. V. GEORGE

NOWADAYS, statistics are influencing our ideas about examinations, and, indirectly, our teaching.

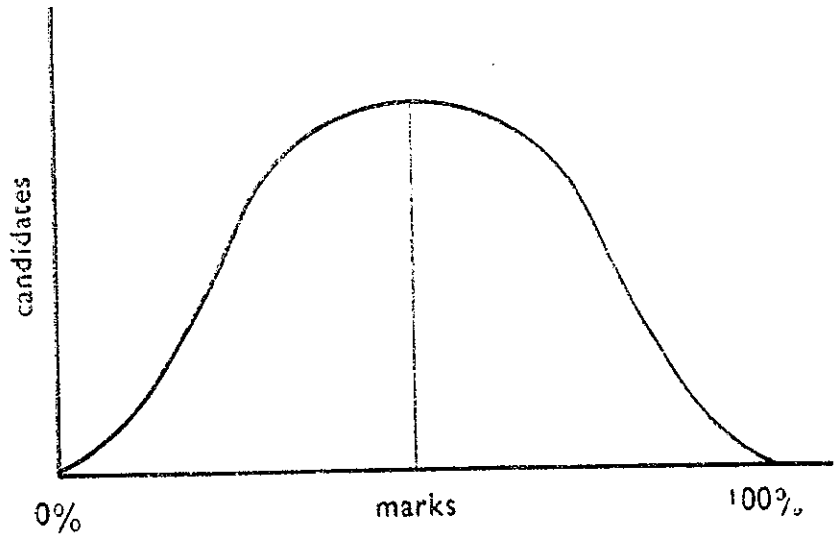
Statisticians have found that the distribution of some feature or other in a population generally follows a pattern to which they attach the word "normal". In graphical form, the distribution appears like this :



A small section of the population shows this feature to a small extent, and a small section shows this feature to a large extent ; and a large section shows this feature to an intermediate extent.

For a statistician, a year's S.S.L.C. candidate is a population, and it is thought that the candidates' marks *should* follow the "normal" distribution pattern, so that we ought to get a results' graph like the graph above, with "number of candidates" in place of "population", and "marks" in place of "feature".

However, as all examination-markers know, there is an embarrassing consequence to a results' curve of this kind. Suppose the pass-mark is at 50%. Now 50% is the flattest part of the curve ; which means that there are many more candidates having marks between 45% and 50% and between 50% and 55% than there are candidates having marks between, say, 25% and 30%. (In some, actual, examinations, the pass-mark is nearer 30% than 50%, but the examination-



markers' embarrassment is not reduced, since in these examinations the flattest part of the curve is around the 30% mark.) As many candidates get one mark fewer than the pass-mark as get one mark above the pass-mark. Almost as many candidates have one mark short of the pass-mark as have the pass-mark itself. Most candidates who fail, fail by only a few marks.

It is obvious that the fate of many candidates depends on one or two marks, one way or the other. Now examination-markers are human, and imperfect, and most people agree that they are not likely to achieve an accuracy, and a consistency of accuracy, to this degree; unless they can be controlled.

For this reason, there has been a great effort to ensure "fair" marking by such control over the markers. It is thought that if unambiguous answers can be elicited, answers which must be either right or wrong, then accurate marking is possible, and the examination can be made "fair". Such examination—"evaluation"—is labelled "objective", to imply that no subjective, or personal, element is allowed to appear. Presumably the ideal is reached when the candidates' work is marked by machines; as is done in some colleges in the U.S.A.

Objective evaluation has, understandably, become fashionable: fairness appeals to everyone, candidates, teachers, parents and markers. Objectivity has become a standard against which examinations are assessed. An examination is "good", according to the consistency with which various examination-markers each give identical marks to the papers of various candidates.

The pursuit of objectivity of this kind is the pursuit of an illusion.

Perfect objectivity—of marking—does not make the slightest difference to the facts inherent in the distribution curve shown above. By objective marking, we are able to state with conviction that 49% was indeed 49%, and that no personal element, in the marking, made it 49% rather than 50% : but this does not alter the fact that almost as many candidates have 49% as have 50% ; this does not alter the fact that most candidates pass and fail by a small margin ; this does not alter the fact that the difference between pass and fail may have nothing to do with difference of merit. A troublesome pen, a dark place in the examination-room, a seat under a fan, a bustling invigilator, any one of a hundred trivial things may decide success or failure. This is a sobering thought—that it may be the state of the point of the nib of the candidate's pen which we are evaluating with objective precision ! More than this, everyone knows that changing any one question in the examination paper would result in a very different distribution—for individual passes and fails—of the candidates on either side of the pass-line. From a statistical point of view, of course, none of these things matter. We have a chance distribution of the elements of chance. Which *individuals* come on one side or the other of the median is of no importance whatsoever.

Which individuals come on one side or the other of the median is, on the contrary, the fact of major importance to the candidates in School Leaving examinations. We conclude that, so long as results produce a "normal" distribution curve, it is an illusion to imagine that objective division of candidates into two groups—pass and fail—is ensured by objective marking.

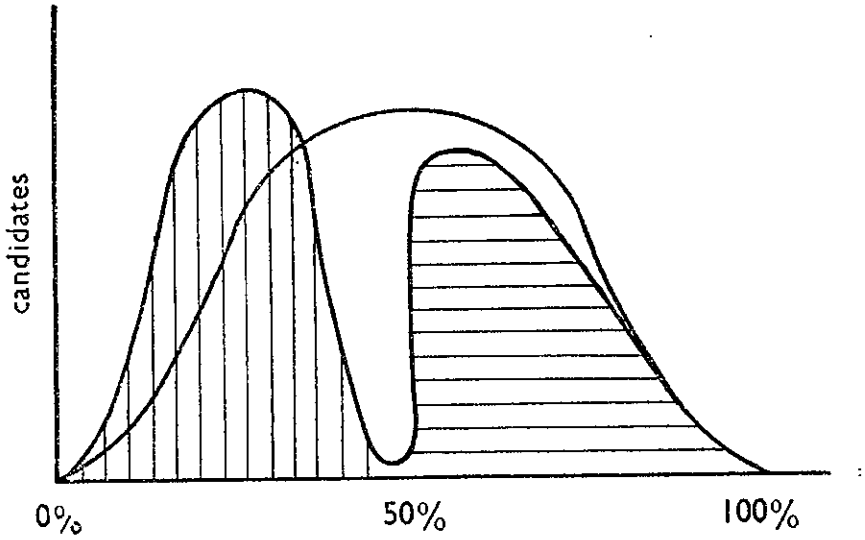
It seems logical, therefore, to abandon, tentatively, the idea of the statistically "normal" distribution curve, and see whether anything can be gained by looking at examinations from another viewpoint.

Surely our difficulty is not with what exams *are* ; it is in what exams *do*. If examinations were required only to measure progress, it would be reasonable to regard the candidates as a population. When we measure features of a population, heights, shapes of head, weights, incomes, ages . . . what we want is an orderly presentation of information. If all we wanted were information, then general progress measurement would be the same kind of measurement as the measurement of height.

However, most examinations are required to divide the candidates into two groups ; that is, our need is not for information, but for decision.

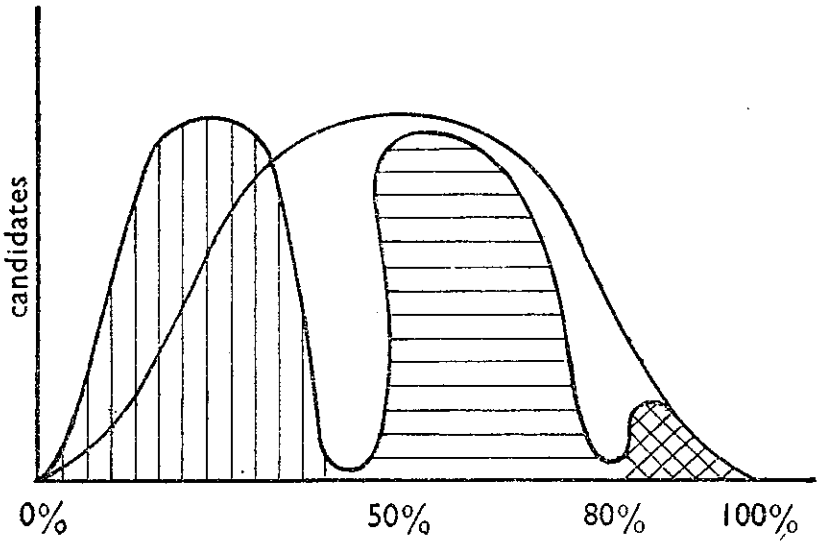
Since this is so, it would seem more sensible to regard the candidates, from the outset, not as one population, but as *two* populations.

The purpose of the examination is the allocation of individual candidates to the one population, or to the other. We want as few candidates as possible to be on the dividing line between the two populations. In other words, if an examination is to decide, its results' curve should be decisive. What we want is:



In terms of marks, if our passmark is 50%, we want our candidates to have either more than 50% or fewer than 40%, either to pass decisively or to fail decisively. If we can get this curve, it is clear that we need not worry about objectivity to an accuracy of ½%. It cannot matter whether a passed candidate had 53% or 58%, or whether a failed candidate had 36% or 37%.

Suppose the examination has a "distinction" section. It may be less important to have a clear decision between pass and distinction than between pass and fail, but it might be desirable. In this case we would want a curve like this:



If we can silence our statistical consciences and preconceptions about "normal" curves (a good way to do this is to conclude that decision by examination is itself "abnormal"), we can now consider how to get curves which allow minimum doubt about the fate of the maximum number of candidates.

In effect, this is not fearfully difficult. The essential is for examination-setters to realise that if they want decision, they must themselves decide. Their decision is: what characteristics are going to determine success and failure. They decide, "We are not going to pass candidates who do . . . such and such definite things. We are not going to pass candidates who do not do . . . such and such definite things." Having made this decision, the examination-setters then see that these deciding points appear in the examination, not once but several times; so that each candidate fails not once but several times, or succeeds not once, but several times on the deciding issues. Similarly, to have a distinction group, there must be distinction-deciding questions.

Examination papers of this kind have something more in their favour: they put responsibility where it should go, on the examination setters. From the teachers' viewpoint, an examination does not just register the accomplishment of the students, or select and reject candidates; an examination determines the classwork leading to the examination, and to a large extent the methods followed in that classwork. In language teaching this is most important to know. If an examination consists of a lot of small, unconnected questions, the teachers' classwork consists of similar fragmentary materials. Whatever work is required in the examination room finds an equivalent in the classroom. It is wrong to think of an examination simply as a measuring device; for the teachers and students the examination decides the programme of work. And from this point of view "objective evaluation" exercises a pernicious influence on language teaching. As you know, in order to satisfy the condition of unambiguous answers, special kinds of question are adopted, regardless of the effect of questions of this kind on classroom teaching. I mean, of course, the marking with ticks and crosses and the striking through of parts of the question material. It is quite wrong, as a classroom technique, to be drawing attention constantly to points of contrast between the foreign language and the mother-tongue usage. It is quite wrong, as a classroom technique, to present incorrect and correct forms simultaneously and exact a choice between them in an artificially isolated context. It is true that a special recognition-skill can be developed, if a teacher is patient enough, so that correct alternatives are chosen—but the success is no indication of what the

candidate does when he does not have the alternatives simultaneously before him. Is this what the examination-setters want? Should the mechanical side of an examination, the technical requirements of the marking, decide what work is done in the classrooms ? Would it not be better for the examination-setters to be quite clear about what specific things should decide acceptance or rejection of candidates, and for them to put these things plainly into the paper, so that teachers follow a programme which has been thought out as a programme ?

To summarise. A statistician's concept of the "normal" distribution of a feature in a population has no validity when we consider examinations, since the candidates are potentially two (or more) populations, not one. Techniques of measurement are only partially relevant, since the purpose of examination is not measurement, but decision, and since objective measurement does not mean objective decision. To divide the candidates into two populations, the examination setters must themselves decide the criteria for success or failure, and multiply the opportunity for success and failure to occur according to those criteria. They must decide these criteria, realising that the teaching programme and methods are being prescribed by the examination.