In our April issue (XVI, 3) we expect to publish two articles on basic problems: 'Language Teaching: Part of a General and Professional Problem' by Professor J. J. Figueroa, of University College of the West Indies, and 'The Structural Syllabus at Work in India' by Professor M. S. Patel, of Baroda University. The 'Commonwealth' series will be represented by C. J. Allen's 'English in Ceylon'. We hope also to welcome back Dr Roger Manvell, writing once more on 'Films from Books and Plays'. Brendan Carroll, Lecturer in Education at Mount Pleasant Training College, Liverpool, has something of interest to say shortly about testing in 'An English Language Survey in West Africa'. Two articles on the English language itself will also appear soon: 'Sequence of Adjectives' by H. Sopher, and 'Notes on the Notional Passive' by M. Beker.

N.B. The attention of prospective contributors is drawn both to the note on our Contents page and to the Editorial of Vol. XVI, No. 1.

ELT Journal (1962) 16, 2: 72–78

Testing—Another Point of View

H. V. GEORGE

A good deal is being written about school and college examinations, and about 'evaluation' in general. This article is about selection examinations, examinations which candidates pass or fail. It is generally agreed that such examinations should do justice both to the candidates and to the subject, and these aspects will be dealt with in turn.

Publicity given to erratic assessment by examination markers has led to a widespread resolve to withhold discretion from them or even to replace them by machines; and a great deal of thought has gone into question-setting techniques to ensure accuracy and consistency of assessment. Accuracy and consistency are summarized as 'objectivity', and it is frequently stated that objectivity lies in the fact that the candidates' marks will be exactly the same whether given by examiner A, B, C, or D.

It is easily assumed that such objectivity guarantees justice to the candidates. However, objective marking does not itself constitute objective testing, for an objectively marked test may be as capricious in its choice of questions, and therefore in its passing and failing of candidates, as any other test. In the second place, there are usually many candidates with just over or just under the pass-mark, and the replacement of any item on the paper by a different though equally valid item would cause a great reshuffling of these candidates around the pass-mark. In other words, *decision*, for them, is arbitrary. In the third place, what puts such candidates on one side rather than the other of the pass-line is probably a combination of trivial factors, a troublesome pen, a seat near a fan, a fussy invigilator . . . It is clear that, when decisions are required, accuracy of measurement cannot by itself ensure their justice.

Various formulas have been proposed for assessing objectivity. I would like to submit the following as the most relevant for a selection examination:

> The objectivity of a selection examination is in inverse ratio to the proportion of candidates placed on either side of the pass-line who, by or but for the intervention of extraneous or local factors, might have been placed on the other side of the pass-line.

To give an example: if, in an examination with a pass-mark of 50 per cent, a large number (say 15 per cent or more) of the candidates have marks between 47 per cent and 52 per cent, then the examination, however marked, cannot claim to be objective for candidates in that group: its deciding function will have been exercised through precise measurement of irrelevant as well as relevant factors.

Justice in decision cannot result from a marking technique; it can be achieved through techniques of decision-making. The fact is that a 'normal' results-curve (a 'bell-shaped' curve) constitutes the worst possible condition for making decisions; for we are using a one-population curve as a device for forming two populations (the passed and the failed). In other words, our use of the curve is incompatible with the nature of the curve itself.

If the examination is to divide the candidates into two groups, we should think of them, from the outset, not as one population, but as two populations. Then the examination-setter determines the characteristics of the 'pass' population, and makes passing depend entirely on the presence of these characteristics. His technique is simple: he involves the decision-making features in several questions,

so that in respect of these features each candidate is either successful several times or unsuccessful several times.

A paper designed to separate the candidates into two groups cannot fail to do so, and the results-curve will be two bell-shapes *with a gap between them*. For instance, in a competitive examination paper set recently by the writer, the best of the 'rejected' candidates had 48 per cent, and the worst of the 'accepted' candidates had 59 per cent. Out of 250 candidates, only one had a mark between 48 per cent and 59 per cent; that is, the examination was doubtful as a selective device to the extent of 0.4 per cent. With a decision-making technique, there is no crying need for extreme 'objectivity' in marking; the examination setter is free to ask questions without the overriding consideration of whether they fit into multiple-choice or other objective test forms.

So much for justice to the candidates. Now we turn to the problem of justice to the subject. Dr R. Lado[1] and others see this primarily as a question of the 'linguistic content' of the test; but checking the linguistic content of individual test-items against items in a syllabus cannot indicate the degree of justice to the subject contributed by the form of the examination itself. To see 'justice to the subject' in this wider context one must realize that the pass-fail examination at the end of a course largely prescribes the teaching during the course.[2]

It is disconcerting to find that, among writers on testing in our subject, the effect of the examination on the teaching is a minor consideration. Mr Gauntlett[3] does not mention it. A working paper at the recent Makerere conference lists seven reasons for conducting English-language tests, but omits the one of first importance: namely, that a purposefully designed examination promotes effective teaching methods.

The effect of the examination on class work represents its justice to the subject in a major sense. There seems little point in deciding the examination pattern, and then adding that 'the introduction of these tests may (!) lead some teachers to angle their teaching in the direction of coaching', or, as Dr Lado does, 'And we will want to ask what effect the test may (!) have on teaching practice.'

[1] R. Lado, English Language Testing, in *E.L.T.*, XIV, 4.

[2] The statement is made explicitly in the University of Birmingham Enquiry into the GCE Advanced Level Syllabuses in Science (Report published Sept. 1959). Many subject panels produced specimen papers intended to alter class teaching. Some panels also stated that the form of the examination affected its efficiency in selection.

[3] J. O. Gauntlett, Education measurement in English, Ehime University, Japan, 1959.

There are several questions to ask. The first concerns the extent to which teachers and students use the language in the classroom. If indeed the examination determines this, then to ensure use of the language in the classroom the candidate must be required to use it in the examination room. Now test items designed for objective marking require the candidate not to use, but to recognize the use of, parts of the language. Advocates of objective testing make a distinction between 'factual' and 'functional' knowledge, and demand examination evidence of functional knowledge. The distinction itself seems entirely reasonable, and one may accept that 'being able to do a thing amounts to the same thing as knowing how to do a thing'.[1] However, one cannot assume that being able to do a thing means in fact doing it. Certainly a foreign student may be trained to select the correct English form when alternatives are presented: this is training in contextual recall of pieces of functional knowledge. No competent practising teacher accepts such recall as evidence that the student habitually *uses* the form he is able to recognize as 'correct'; and if recognition of the 'correct' form from alternatives means examination success, then the student's own use of the language becomes irrelevant in classroom preparation for the examination. All multiple-choice questions, whatever their linguistic content, are in this respect unjust to the language.

In the writer's opinion, functional knowledge should not be examined in this way until an advanced stage; until students use English easily there is no place for choice from among alternative forms. In the very early stages the pupil's use of English is automatically restricted and there is no need for minute control of individual items. In the middle stages, there are established ways of control: for instance, by a list of questions, or by key words and phrases around which a narrative or argument is developed. Later, the following type of 'writing to prescription' is possible:

> Write a short account which begins by stating a habit in the past, then states the occurrence of an event causing the replacement of that habit by a present habit, then expresses pleasure at the result, and hope of its continuance in the future.

At each stage the constructions the candidates use can be controlled without their being prevented from using the language.

The second question concerns the effect of following the maxim: 'Tests must deal with one category of skills at a time.' When a pupil

[1] J. O. Gauntlett, p. 9.

gets to the stage of using the language, even in the first few weeks, he uses skills in combination, not isolation; and many teachers fear the recent isolating and fragmenting tendency. It is not the separate *marking* of spelling, grammar and so on which matters, but the use of specially contrived, problem-isolating sentences for convenience of marking: this must have a pernicious effect in the classroom.

The third question is whether the examination encourages discretion. It is useless for a teacher to stress the precepts 'If you don't know it, don't use it,' and 'There's always another way of saying it,' unless the examination permits discretion. The testing theory evolving from objective test patterns is against discretion. For instance, criticizing written composition, Dr Lado writes: '. . . the student is able to avoid those (problems) that trouble him, which would be the ones we would like him to attempt.' An exact picture of his imperfections may be unobtainable, but we have a better picture of what a student *can* do if we allow him to avoid what he cannot do.

The fourth question is whether the foreign language is to acquire a validity of its own, or be based on a comparison between foreign language and mother tongue for each point of expression. Testing, Dr Lado says, should be based on language areas where interference from the mother tongue is greatest: 'From this we simplify our view of what it is to learn a foreign language, by seeing that it is the mastery of the difficult elements and patterns.' It seems to the writer that the degree of divergence between mother tongue and foreign language is an eccentric qualification for classroom attention. It is more usual, surely, to determine the amount of classroom attention according to the degree of usefulness (whether it diverges or not) which an element or pattern has in the language.

This is a very important methodological consideration. If, under examination influence, we apportion class time according to the difficulty of each word or structure, our students' English is going to be very different from standard English; since the relative importance of individual structures in native English usage is not represented by their relative importance measured according to their difficulty for a foreign learner. One of our hardest tasks is to persuade students that when they have made the effort to learn a word, an idiom or a structure, that effort is not, in itself, an authorization to use it. The Lado 'new view' of learning a foreign language is not really new. English learnt by 'mastery of the difficult elements and patterns' is what we call 'Ausländer Englisch'.

Fragmentation of language material inhibits the development of a feeling for the relative importance of word or pattern, and promotes

indiscriminate attention to detail. This is no theoretical objection. Dr Lado exemplifies the statement explicitly with his 'important problem in auditory comprehension', distinguishing between 'washing' and 'watching'. Most teachers are happy if their students understand an Englishman speaking normally, and if an Englishman understands them. In normal speech contexts, failure to distinguish between /ʃ/ and /tʃ/ does *not* affect aural comprehension; comprehension of separate phonemes is not a skill distinguishable from the total skills of 'comprehension'. Naturally it is inelegant if our students say 'wash' for 'watch'; but only someone habituated to regard a language as the sum of a number of equally essential fragments could believe that this pronunciation problem has a significant connection with 'auditory comprehension'. A student saying 'wash' for 'watch' has created a problem of no more significance than the thousands which the language supports quite happily, for instance the one created if I ask the headmaster, 'May I have that form, please?' when 'form' may be a piece of paper, a bench, or a class. This problem, like Dr Lado's, is a problem only when the context is concealed. What Dr Lado suggests is that we *invent* confusing contexts, for testing, which would have little chance of occurring in real life. Whatever their defects, dictations are nearer to real life situations than these phonological problems, which are most unsuitable as examination material.

So far from reality are testing techniques leading foreign-language work that Dr Lado can write:

> We are thus able to break away from having to ask the student to speak when we test his ability to speak, since this procedure is inaccurate and uneconomical . . .

If, to show their ability to write, our students do not have to write, and to show their ability to speak they do not have to speak, we are indeed beginning a new era in foreign-language work.

I have tried to show that justice in selection examinations is not attainable through objective marking techniques; but that it can be achieved through examination-paper setting with clear criteria for the acceptance and rejection of candidates. Then I have urged that the main importance of selection examinations is their function of determining course content and teaching method; and I have urged rejection of objective-type testing in language examinations of this kind, since it promotes undesirable practices:

(i) training in contextual recall of facts of language behaviour (instead of sound language habits);

(ii) undue isolation and fragmentation of skills;
(iii) discouragement of discretion in the students' use of the language;
(iv) the methodologically unsound application of time and attention according to learning difficulty, instead of according to usefulness (as indicated by native-speaker's usage);
(v) the contrivance of artificial question-material, presenting 'problems' individually insignificant.

I have tried to indicate that sufficient examination control over the candidates' use of the language may be effected without the promotion of undesirable practices.

Incidental Classroom English

A. S. HORNBY

(*Mr Hornby is the author of* The Teaching of Structural Words and Sentence Patterns *and many other works on English and English-teaching*)

Most teachers of English are aware of the value of using the language for occasional incidental comments and requests, outside the routine question-and-answer drills, explanations of difficulties, and other normal procedures. They are aware of the value of encouraging their pupils to use English in similar ways. They know that it is wise to keep the use of the mother tongue to a minimum, and to restrict it to explanations and descriptions where the use of English would be uneconomical of time and perhaps inadequate for the purpose.

Some teachers are understandably doubtful about the kind of English to be used. Are they to limit it to such simple imperatives as 'Open (close) your books', 'Hands up those who know the answer', 'Take up your pens', 'Come to the blackboard', and an occasional comment such as 'Very good' or 'Excellent'? Or are they to use English freely whenever opportunities occur? Some teachers look for lists of commands, requests, and formulas for classroom use. What, they ask, do teachers in Great Britain say when they give instructions to a class during a routine lesson?

There can be no satisfactory answer to this last question. Much depends upon the relations between teacher and pupil, upon whether