EDUCATION IN THE COMMONWEALTH

EXAMINATIONS AT SECONDARY LEVEL

COMMONWEALTH SECRETARIAT

EDUCATION IN THE COMMONWEALTH

EXAMINATIONS AT SECONDARY LEVEL

Commonwealth Secretariat

Published by the
COMMONWEALTH SECRETARIAT

To be purchased from the
Commonwealth Secretariat
Printing Section
Marlborough House
London, S.W.1.

S.B.N. 85092 034 5

Education in the Commonwealth

Foreword by Mr. Arnold Smith,
Commonwealth Secretary-General

EXAMINATIONS AT SECONDARY LEVEL

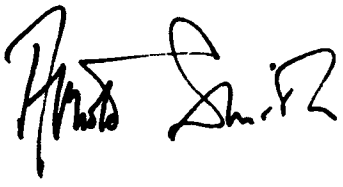
FOREWORD

The purpose of this series of occasional papers, EDUCATION IN THE COMMONWEALTH, is to extend the service already provided by the Commonwealth Secretariat through its conferences, seminars, reports, research, correspondence and visits.

By drawing together material on selected topics of wide educational interest it is hoped that the Secretariat will further promote the interchange of ideas and information among individuals, institutions and countries.

Examinations at the Secondary Level, the theme for the First in this new series, reviews an area of pan-Commonwealth interest. So many of the efforts in the development of new curricula are eventually frustrated by established examination systems. While many have been critical of traditional examinations, the new methods, too, have not been received without scepticism. Meanwhile the demand for educational yardsticks must be met. Equally important as the measurement of the child is the measurement of the system. Every endeavour to make suitable secondary education available to greater numbers of children brings with it the need for close scrutiny of the nature, purpose and goals of that education, and of how well they are achieved. Evaluation involves assessing the product of the system, and when attempting to evaluate, the necessity for examinations confronts us yet again.

Throughout our member countries, and indeed throughout the world, education is being reconsidered, restructured and replanned. The search for increased effectiveness, quality and relevance results ever more frequently in a greater national orientation of the educational system. So it should, but it does not follow that every educationist need carve a painful path in isolation. Much saving of time, effort and already scarce finances can accrue from a knowledge of the techniques and strategies investigated and subsequently adopted or even discarded elsewhere. Failure can be no less instructive than success! Something of value can be derived from every investigation, and in presenting this collection of articles it is hoped that new possibilities may be evoked and new directions indicated.



Introduction

Examinations, testing and measurement, assessment and evaluation, play an essential role in education. They also attract more emotion on the general public than any other aspect of education. This is understandable, for the parent sees his own adequacy under scrutiny as his child is tested and measured, selected and allocated. For their part, the testers, from classroom teacher to regional examining board, are concerned with the construction of valid, reliable and appropriate instruments from which conclusions can be drawn with a fair degree of certainty.

The road to the examination hall is paved with good intentions, and some consolation for the inadequacy of examination techniques developed so far may be drawn from an objective consideration of the consequences which could arise were they ever perfected: "That will be a sorry day for human initiative and personal aspiration on which there is announced the examination which shall be completely valid, the perfect predictor of each individual's ineluctable future."

The first two articles published here, by McIntosh and Matys, approach the general question of examining from two distinct directions: the first anxious that examinations should not seek to exceed their rightful function, the second outlining methods and techniques suitable for different purposes.

Because in many countries the demand for particular forms of education exceeds the supply, competition for the available places results in a distortion of the teaching-learning situation in the period prior to the selection examination. The great majority who are not selected thus suffer for the benefit of the few who are. From the administrative viewpoint, of course, the smaller the percentage of successful candidates to be selected the less efficient the selection process need be: "It is sometimes urged that, while errors in selection may be inevitable, few of them are serious. The good candidate, the argument runs, would be selected by almost any method. In that more critical region where there seems to be nothing to choose between candidates, and where they are most numerous, no grave errors can be committed whatever decision is made."

Criticisms are increasingly levelled at the discriminatory effect of many examinations. "'O' levels might be described as the epitome of bourgeois activity throughout history. They are the rewards for clerkly diligence." Pilliner's article shows how difficult it is to construct tests which are "culture fair", affording a real equality of opportunity for all the students taking them. This is an immediate problem in many developing countries where the rural child tends to have less success than the urban child in the formal school situation. This widespread problem is taken up again by Taylor, drawing on his experience in the Pacific region.

The procedures for test construction described by Elley represent an encouraging step towards the use with small groups in the classroom of objective-type tests designed and produced by the teachers themselves.

Hitchman's review of the examining of Spoken English, probably the most subjective form of examining, explains means by which an acceptable degree of standardisation can be achieved among examiners, each of whom will usually be working on his own. The practical examining of science raises problems of a different kind, as Dave and Patwardhan indicate. The development of valid and reliable practical examinations is attracting much attention as educationists attempt to identify means by which the knowledge and skills of large numbers of examinees can be suitably tested. In particular, the provision of adequate supplies of equipment for this purpose is impossible for many schools in developing countries.

Two articles outline the administration of examinations over extended geographical areas, the South Pacific and East Africa. The Fourth Commonwealth Education Conference, meeting at Lagos, Nigeria, in February 1968, noted the successful operation of co-operative arrangements for examining existing among certain regional groups of the Commonwealth, and it would seem that there is scope for more activities on these lines. Certain examinations can be administered economically and efficiently on a very large scale, although Kiwanuka's description of the establishment of a new regional examining body indicates the many obstacles which have to be overcome if the organisation is to run smoothly.

Guidance and counselling are relative newcomers to Commonwealth secondary schools; Iyer's account of Malaysia's pioneering effort in this field may provide a timely stimulus to other countries to consider the introduction into their schools of a similar service.

Secondary school examinations have long been used for two distinct purposes: to assess attainment at the end of a course of secondary education, and to determine the potential of an applicant for third-level courses. Recent researches confirm the low correlation between final secondary school examination results and success at the university: "Since the performance at A-level in the subject of the Honours School has little or no prognostic value, considerably less weight should be attached to it in the future." Interviewing adds little of value to the selection procedures. In these circumstances educationists are turning to tests specially designed for candidates wishing to proceed to third-level courses; Traub and Elliott explain the operation of one of these , the Canadian Scholastic Aptitude Test.

The last two articles deal with factors affecting candidates' performances. Mackay's investigations indicate that in essay-type examinations with a choice of questions, the poorer candidates tend to choose the more difficult questions, thus minimising their achievement. The inclusion wherever possible in an examination of a compulsory section, possibly comprising questions of an objective type, would make possible direct comparisons between candidates. It would also provide useful indicators of a candidate's not doing himself justice in the optional sections of the examination because of an unwise choice.

Anxiety and stress are commonly believed to exert a powerful influence on examination performance. Sinclair's investigation among Australian High School students of above-average ability show that in some cases a degree of stress may serve to raise the performance of some individuals, although means must be found to reduce the adverse effects on "high anxious" candidates.

These articles provide only a brief survey of some of the problems of examinations at secondary level common to wide areas of the Commonwealth. No mention has been made of the many valuable experiments in new methods of examining, such as "open book" examinations, the use of continuous assessment, or the compilation of student profiles. Nor has the interrelationship of curriculum and examination been adequately covered. A list of research projects on examinations is, however, reprinted from "Education in the Developing Countries of the Commonwealth: Abstracts of Current Research 1969", published by the Commonwealth Secretariat.

This small collection, produced by a dozen contributors in ten Commonwealth countries in five continents, is published in the hope that it will stimulate exchanges of experience and information among those working in this field. So often does it prove that the activities of researchers overlap. If this Paper encourages interchange, by direct correspondence or through the Commonwealth Secretariat, it will have achieved its purpose of helping to save time, money and expertise, all of which are scarce commodities.

EXAMINATIONS

Douglas M. McIntosh

Principal, Moray House College of Education, Edinburgh, Scotland.

Examinations have a determining influence on careers, and accordingly they arouse emotional attitudes: they have been referred to both as "the central nervous system of education" and as "a millstone round the necks of the schools." Such reactions arise partly because examination results are used without sufficient thought to their validity and partly because many examinations are not constructed with clearly defined objectives. Despite the early researches which revealed the extent of the unreliability of examination marks, only recently have steps been taken to produce the improvements which are so necessary in any system of large-scale examining. For example, the pass/fail examination is a weak measuring instrument because all measurements of human ability are approximate and to use an examination to distinguish between a pupil who scores 50 and is awarded a "pass" and another who scores 49 and is awarded a "fail" assumes a degree of accuracy which is beyond the resources of a written examination.

A very large subjective element enters all aspects of examinations. The selection of questions involves the sampling of sections of a syllabus on which the examinee is tested: the sampling depends entirely on the views of the examiner. Pupils have sat examinations in which they have "spotted" certain questions and, accordingly, found the examination easy, whereas others have had the experience in which much of their hard work has gone for naught because the sample has been drawn from areas of the syllabus on which they had not concentrated their attention. A subjective element is also involved in the allocation of marks. For example, a group of teachers or students when given no guidance will award a wide variety of marks to an arithmetic question in which the pupil takes down one figure wrongly but works the sum out correctly: the different marks indicate different value judgements.

Examinations have a wide range of purposes. The following are the most common:

(a) attainment : standards and objectives must be clearly defined and the examinees should be given some indication of the standards which they are expected to attain.

(b) diagnosis : such tests or examinations attempt to find out what the pupil does not know. For example, a diagnostic test in arithmetic may show that a pupil has a specific weakness: many young children have difficulty with zeros and when such weaknesses are revealed by a specially constructed test they can be corrected.

(c) prediction : most of the examinations at the end of secondary school are used to predict the success with which a pupil will engage in a course of higher education. Many such examinations have not been specially constructed for this

objective, and it is merely assumed that they have this value although investigations have shown that the relationship between success in higher education and secondary shcool examination results is not high.

(d) motivation : the knowledge that a specific standard has to be achieved by a certain time gives purpose to teaching and to learning. Teachers interested in their pupils who are attempting to gain admission to university will strive hard to ensure that the pupils gain the necessary entrance qualifications, and the pupils for their part have a clearly defined goal which channels their energies.

(e) selection : selecting a number of pupils for a particular course of study, where the purpose of the examination is to ensure that pupils who are chosen have the ability and the knowledge which will enable them to undertake a particular further course of study. When a restricted number of candidates have to be selected from a group, it is assumed that the examination can make the fine distinction between the candidates at the border line.

(f) a teaching instrument : too often, once the examination marks are issued, no further work is done on what has been written. On the other hand, examinations or tests of some nature should be a regular feature of classroom work. These need not be formal examinations but tests to check whether the learning experiences created by the teacher have been effective. In all teaching situations, the teachers should have clearly defined objectives. Following the exposure of the taught to a learning experience the teacher must attempt to assess the success with which the objectives have been achieved. The assessment will have two purposes: first, to ensure that the learning experience is effective, and secondly, to gain some insight into the ability and attainments of the taught.

One of the weaknesses of examinations is that they tend to be used for a wide variety of purposes. Examinations at the end of secondary school courses are used for entrance qualifications to a large variety of professions, which select these examinations merely because they are available. It has often been suggested that different professions should set their own examinations, but this would place the schools in an impossible situation whereby they would require to prepare students for a very wide variety of examinations. National external examinations such as the GCE in England and the SCE in Scotland are used for entrance qualifications for many courses in higher education and while they were orginally intended for university entrance they now are used by a multiplicity of institutions providing higher education: it is doubtful whether they can have a high predictive value for all institutions.

An examination should test the objectives of a specified course of study, and these objectives should be clearly defined. In the case of schools, teachers should create the learning experiences which enable pupils to attain the stated objectives and the examination should provide a measure of the effectiveness of the learning experience. The knowledge of facts is essential in any course, but too many examinations rely heavily on memorisation. In an examination on statistics or mathematics pupils are allowed to use mathematical tables, but there seems no good reason why they

should not also be allowed any books which they use in the classroom. The memorisation of a formula is of little importance in life outside the shcool: it is the ability to use the formula which is of real significance.

Objectives have been classified by Bloom in his widely known work "Taxonomy of Educational Objectives"* and it is possible to analyse examinations showing the percentage of each category which have been tested by the examinations. The categories are:

(a) information
(b) understanding
(c) application
(d) analysis
(e) synthesis
(f) evaluation

These are in what is known as a hierarchical order: each category cannot be carried out without skill in the previous categories, for example, understanding is not possible without information and, similarly, application cannot be carried out without information and, similarly, application cannot be carried out without information and understanding.

Too great reliance is often placed on examination marks: as has been indicated the pass/fail examination provides an example of an assumption of the accuracy which it is assumed examinations posses. Much research has been conducted to show the unreliability of examination marks. Pupils sitting similar examinations at intervals often show a degree of unreliability which is difficult to understand. The lack of consistency in examination marks may be due to three factors:

(a) the unreliability of the marking: research has shown that the same examiner can award different marks to the same pupils after an interval of time. In one investigation 14 examiners were asked to re-mark 15 history scripts some 12 to 19 months after the first time, having kept no record of their previous marks. The examiners awarded not only numerical marks but also the verdict of failure or pass or credit. It was found that in 92 cases out of 210 the individual examiners gave a different verdict on the second occasion from the verdict awarded on the first occasion.

(b) the unreliability of the examination: questions in any examination can deal with only a sample of the syllabus which has been studied. If the sample is not representative of the syllabus as a whole the examinees may not do themselves justice. When a teacher can make an accurate forecast of the type of question likely to appear in an examination, his pupils will probably gain marks unrepresentative of their ability. To give consistent results an examination should be of the same order of difficulty to all examinees of comparable ability.

---

* Bloom, B.S. (ed): Taxonomy of Educational Objectives, Handbook I: Cognitive Domain, David McKay & Co., Inc., New York, 1956.

(c) the unreliability of the examinee: few individuals can consistently reach the same standard in any activity: all have their good and bad days. Some are more consistent than others and it is for this reason that a decision on a pupil's future should not be based on the result of a single examination. Ill health, or nervous tension, may also cause an examinee to do badly in an examination.

One technique which has been successfully developed to overcome the unreliability of examinations is the objective test. "The Battle of Bannockburn was fought in - 1413 1143 1314 1134 - underline the correct answer." Questions of this type have only one correct answer and thus the mark is likely to be the same no matter who corrects the paper. A large number of such questions make up a test and therefore a much wider sample of the syllabus can be studied. Such tests have some obvious disadvantages in that they cannot test a pupil's ability to select data and arrange it in an orderly and logical order. The construction of such tests also is a much more laborious and time consuming operation than an essay type of examination.

The most important factor in determing success or failure in examinations is motivation. Where pupils have a strong desire to succeed they are more likely to do well: where they have little interest in the result the standard achieved will be low. Care has to be taken, however, to make sure that there is no over-anxiety on the part of the examinees, otherwise they become upset and do themselves less than justice. One of the causes of uneven performance by examinees is the undue propo'rtion of an examination dependent on memorisation: too many examinations cannot be answered without remembering information. For example, the following question is typical of some examinations - "How many balls of $1\frac{3}{4}$" radius can be made from 539 cubic inches of metal?" Examinees may be capable of answering this question, but cannot remember the formula for the volume of a sphere. Outside the examination room, anyone who had forgotten such a formula would simply turn to a book for the answer. The present explosion of knowledge makes memorisation more and more unprofitable since much information becomes rapidly out of date. To find out how to use facts is much more important than remembering them, and examinations should reflect this change in emphasis.

It cannot be too strongly emphasised that all assessment of human ability is approximate, hence the more reliable and valid information about an individual which can be ingathered, the more any assessment is likely to give a true measure of the individual's ability or attainment. The result of a single examination should never be used as a means of assessment. Three components ought to be used:

(a) teachers' assessments : the pupil's record over a number of years will give a rating which should be fairly reliable. Teachers standards, however, vary according to the teacher's experience, hence some form of external measure is necessary to scale the assessments on a uniform standard.

(b) external tests of objective type : these give a reliable measure of a pupil's ability in a clearly defined syllabus.

(c) essay type examinations : these measure other aspects of achievement and ability than those generally measured by the objective test.

Examinations are not the same thing as the day of judgement: and they are certainly not the Alpha and Omega of education, nevertheless some form of assessment is essential if teaching is to be effective and the potential ability of individuals is to be developed: the ruler is unlikely to be discarded because it cannot measure to 1/1000th of an inch.

TESTS AND MEASUREMENT PROCEDURES, REVIEW AND EVALUATION.*

G.J. Matys
Curriculum Research and Development Unit
(Measurement and Evaluation Section),
Ministry of Education, Ghana.

PART I

TESTS AND MEASUREMENT PROCEDURES:

To a considerable extent modern education is characterised by the emphasis it places on adapting the educational programme to the needs of the individual child. Since these needs are governed by the child's level of ability and by the degree to which he has mastered the educational contents to which he has previously been exposed, it is important to determine these factors as accurately as possible. Once this information has been obtained, much can be done to individualize the educational process for each child, in part by grouping children into homogeneous instructional groups and in part by differentiating instruction within the classroom. It follows then that accurate educational measurement is a prime and key factor in modern educational trends.(1)

So if in modern education the emphasis is on teaching the individual it follows that we must have knowledge of, and differentiate between, individuals. Knowing the individual requires evaluating and testing. At the same time everywhere in the world the role of testing and evaluation in education is being questioned, criticized and scrutinized as never before.

Now before one goes further into the subject a definition or two should be made. The first is that Educational Evaluation is much broader than "testing". Educational Evaluation uses a variety of methods to measure and assess. These include questionnaires, surveys, cumulative records, projects, class work, oral answers, role playing and so on. Secondly the tools of educational evaluation are not a precise measure as are the tools of the engineer and physical scientist. In education one is trying to measure aptitudes, or intelligence, or content and processes - all very intangible and very difficult to assess. So it must be taken as a premise that the best of tests, under the best of conditions, provides large factors of error. When conditions (the training of personnel, administration and the tests themselves) are not ideal, results are even less reliable and less meaningful.

Testing had a traditional and fixed role for many years. It marked the end of one phase and the beginning of another. It meant passing out of one grade and into another. Successful completion of examinations allowed one to enter a profession, such as medicine. This type of testing has something in common with the initiation ceremonies characteristic of many non-Western societies and of various secret or exclusive groups within Western society. The examination or the initiation ceremony is a more or less

---

(1) Test Service Notebook - Test Bulletin, Harcourt Brace and World Inc.

* Originally included in the documentation for the Commonwealth Conference on Education in Rural Areas, held at the University of Ghana, Legon, Accra, Ghana, 23 March to 2 April 1970.

difficult procedure; if the examinee reaches a certain level of performance, a level agreed upon by the elders of the society, then his status becomes altered and he becomes permitted to practice as a doctor: the high school leaver is enabled to seek employment. The characteristic of this type of examination is that it marks the end of one phase in the person's life, and it demonstrates that he is competent to enter a new phase. It is perhaps worth making another distinction between (a) the terminal examination in a particular vocation, which is intended as one indicator of the individual's competence (and the examination should not be the only indication which is used), and (b) the final examination at the end of a non-vocational course. The former should ensure that society is not plagued with incompetent doctors and other professionals, but the latter has a less clear social raison-d'etre, and would appear to be a potent agent in the development of a society in which every adult's status is determined by his scores on tests taken during his adolescence.

The second reason advanced for examining is as part of a continuous process of education, a method by which the teacher assesses what each student has and has not learned. This use of examinations has been going on for many years and it might seem hardly to deserve comment, - but the present renewed interest in "Measurement and Evaluation" seems to stem from a more precise analysis of the assessment process than used to be practised.

In modern education this first type of testing plays a less and less important role. Testing now is seen as one type of evaluative procedure and only a part of the overall educational process. There was a period recently when some advanced countries had an almost religious faith in tests, whether they were tests of aptitude, ability, achievement or intelligence. Test results were felt to provide final, definite, and reliable answers to many questions. In another, earlier, period educators ignored and discounted tests and their results as useless. It was felt that one could not measure "intelligence" and other "intangible" processes of man.

Today a middle-of-the-road approach which avoids either of the above extremes is gradually emerging. Educators are realising that tests are far from useless, and yet far from providing all absolute answers. It is realised that tests are only one factor, one piece of information which becomes valuable when combined with school marks, common-sense evaluations of teachers and a multitude of other data, some scientific, some less so. Also it must be kept in mind that measurements are only tools, a means to an end, and not an end in itself. We do not weigh or measure an article just for the sake of knowing how heavy or how long it is. We use this knowledge in some way. So it is with testing - there must be some definite purpose in the testing we do. Are the tests simply to measure achievement? Are they diagnostic tests only, to be used to diagnose teaching weaknesses and learning difficulties? Are the tests to measure a person's potential for academic or other fields, his capacity or aptitude? Do the tests try to measure interest, attitude, intelligence or personality? There are instruments available today which attempt each of these, or some combinations of these, and so the purpose must be clear from the start.

Generally the types of tests listed above are Standardized Tests - tests built by experts over long periods of time and with carefully selected norms and so on. However, both these and teacher-constructed tests have a role or purpose in the classroom. What are some of the purposes of classroom testing?

Let us examine some of the purposes and uses of such testing (this applies whether the tests are "teacher constructed" or "standardized"):

(1) To test pupils' achievement. This is probably the most common purpose of testing. The teacher should have constant feed-back on how well the skills taught have been mastered and how well the concepts and understandings can be applied. However, the diagnostic aspects of achievement tests should not be overlooked at any time.

(2) To assess the effectiveness of instruction. Educators are so prone to say, when looking at the results of a test, that the pupils have done "well" or "poorly". Frequently the results of a test are more an indication of how well the teaching has been done. If the results of a test show weaknesses, the teacher has the opportunity to re-teach, change the method of approach, or seek for other methods of increasing the effectiveness of instruction.

(3) To motivate pupils to improve in their work. Test results will encourage most pupils to put forth their best efforts. A word of caution is in order here. Every class has pupils of varying abilities. We do not expect all of them to run at the same pace when they are racing. In the same way it would be wrong to expect all pupils, the bright and slow ones, to achieve the same standards on a test. It would be wrong, therefore, to compare the mark of a pupil with lower ability with that of a pupil with higher ability. It is sound, however, to stimulate pupils to improve their own marks on successive tests rather than comparing them with the brightest pupils.

(4) To discover individual problems and weaknesses. The test results will identify pupils who have particular problems and the teacher then has the opportunity to provide individual help and instruction to such pupils.

(5) To provide a sound basis for keeping parents informed regarding pupils' progress. Parents are usually interested in knowing how their children are performing. If full records of test results are kept by the teacher, these form a good basis of communicating pupil progress to the parents.

(6) To locate or identify weak areas in the teaching-learning situation. If test results are analysed carefully "gaps" or weaknesses may be discovered and necessary steps taken to deal with them. Methods of doing this are mentioned later.

(7) To gain information for grouping pupils for instructional purposes. It has already been mentioned that pupils within a class vary greatly in their innate ability to learn. The slower pupils require simpler explanations and more instruction and drill. This frequently becomes boring to the brighter ones and causes them to lose interest. Test results would indicate which pupils might be grouped to provide the most suitable instruction.

(8) To gain knowledge about individual pupils for guidance purposes. If full records of test results are kept pupils' strengths and weaknesses as well as their special interests will be discovered. This information can be used in guiding pupils to make proper choices when they go on to further education or when choosing a vocation.

The above suggestions apply particularly to teacher-made classroom tests. Standardized Tests that are prepared for a wider use, such as throughout a school system or country, would serve other purposes as well. Such tests should provide even more help and information to teachers, parents, administrators, curriculum makers and educational planners and policy makers. This additional information would help:

(1) to evaluate courses or syllabuses for the purposes of revision, etc.;

(2) to compare different methods of instruction and assess teaching methods;

(3) to ascertain standards of classes within a school, a district, a region or the country as a whole;

(4) to assess the work of individual teachers;

(5) to assess pupils and recognise the individual characteristics of each pupil (we should know: (i) his difficulties and weaknesses
(ii) his strength and present knowledge);

(6) to provide a means for pupil and teacher review, an integral aspect of learning;

(7) to provide information for educational planning and policy making.

SOME OTHER GUIDELINES

It is worth emphasizing that testing should be done only when we know how and by whom the results will be used. Of all the functions for which tests may be used, the least valuable function educationally a test can perform is when it is used only by administrators and only in passing, failing, admitting or screening students. Yet examinations with this function alone are still quite common. The most valuable function of tests is in helping the pupil and teacher communicate, and derive benefits from the learning process. In addition tests should be constructed with a clear knowledge of all their educational objectives and should be critically evaluated to see whether they are valid (measure what they are supposed to measure) and are reliable (consistently measure the same thing in the same way). The interpretation of course must be logical, attributing no more, or less, to a particular result than it deserves.

It is worth noting also that if testing and evaluation are to become an integral part of the educational system, teachers must know something about the field. Generally today teachers learn how to demonstrate, explain and put a point across. But little or nothing is given them on how to evaluate,

get feedback and measure what changes have taken place in the pupil. And yet to be truly effective a teacher must know (a) what the pupil has already; (b) what he has failed to learn and, if possible, why; (c) what the pupil is capable of learning.

There are, then, certain problems and pre-conditions to good testing that must be attacked simultaneously with any attempt to enlarge the role of evaluation in schools. These include:

(1) A FOUNDATION OF TEACHER TRAINING in the understanding, interpretation and use of tests. Programmes must be developed to improve this in the colleges and through in-service training. A basic record-keeping system (cumulative records) is needed together with teachers who can use it properly.

(2) DEFINED OBJECTIVES. Good evaluation programmes can help to show how far the school programme meets the objectives of education. This presupposes that there are measureable objectives set forth for general education as a whole and also detailed objectives for each subject. It is only against some objectives, however simple, that one can evaluate.

(3) OTHER GENERAL PROBLEMS TO TEST DEVELOPMENT IN A DEVELOPING COUNTRY.

(a) Difficulty with control groups due to seemingly high turnover among pupils and teachers, very different levels of teacher training, lack of records of ages and other data etc.

(b) Lack of pupil and teacher familiarity with the notion of (i) carefully timed tests (ii) objective tests.

(c) Greater differences than in developed countries between urban and rural cultural factors.

(d) Administration problems (developing the effective machinery necessary).

(e) Language problems. Literature and other evidence suggests that any standardized test meant to measure "anything". In countries where English is not the first language it will, in fact, measure largely facility with the English language.

THE INFLUENCE OF TESTS

In a modern technological world there is bound to be a great concern with accurate measurement. Scientists can calculate an exact point and time for a moon landing 240,000 miles away. It is inevitable, then, that this desire to evaluate accurately should spill over into education. And, as is pointed out elsewhere in this paper, the two extremes (a) of attributing too much, and (b) too little, significance to the role and value of testing, both exist.

One author says "Measurement touches upon and influences every phase of education. Whether it is marking, promotion, guidance and

counselling, curriculum development, instruction or some other aspect of the work, measurement plays an important part."(2) Examinations and marks can be called the currency of education. By these marks, or value assigned, people are passed, granted certificates, promoted, given degrees and so on. We often judge a man's worth by his academic percentages!

There is general agreement then that testing can and does have a profound effect on the educational system of a country. The methods of teaching, the emphasis in the curriculum, the attitude of teachers and students, are all affected or sometimes dominated by the examinations. The types of things stressed in examinations largely determine what happens in the classroom. It matters little what teaching notes or syllabuses are prepared unless the examinations reflect the same spirit and aims. This is especially true where there are large scale and important external examinations.

EXAMINATIONS

Although examinations should measure what is being taught in the classroom, it is very easy for the situation to develop where we teach what is tested rather than test what is taught. Curriculum development and examinations cannot be separated and should be developed in close harmony at all points. Persons sitting on Curriculum Panels or Examination Panels should both be familiar with the general national aims and objectives of education as well as the specific spirit and aims of a given sullabus.

All this does not mean that examinations are the only determining factor in education nor is this a criticism of external examinations. The important thing is that these things should be in the right order and priority; tests should serve the educational goals and needs, not determine them.

The author recently sat on a committee the members of which were drawn from the Ghana Ministry of Education and the West African Examinations Council. A paper produced as a result of these meetings had in part this comment on examinations:

"In spite of inherent weaknesses external examinations are useful and necessary in many situations. In Ghana, for instance, some common measure is needed to provide objective norms and maintain a common standard owing to great disparities in:

i) staff;

ii) training facilities;

iii) libraries and supply of textbooks, etc."

The point here, then, is not to weigh the advantages and disadvantages of external examinations not to debate how much external examinations can affect classroom practice, for such argument or debate has limited value.

---

(2) V.H. Holl, Introduction to Educational Measurement, Haughton Mifflin Co., Boston, 1965.

INTEGRATING EXAMINATIONS AND CURRICULUM

The important thing, then, is to recognise

(1) that there is interaction between examinations and curriculum;

(2) that examinations are not simply passive instruments of assessment but an integral and vital part of the educational process;

(3) that both examinations and curriculum are important and powerful forces for change;

(4) that both form part and parcel of the educational process; and

(5) that both should be under constant review in terms of relevance to changing needs.

The central problem - which must be true of every educational system - is to find the most effective ways of ensuring that curriculum planning and examinations complement each other and work towards the same end. In other words, what should be done is to make sure that the examinations used (a) reflect the same goals (b) promote the same spirit, objectives, emphasis and priorities that the curriculum planners had in mind. Without the proper integration with curriculum, a tester starting from the same written syllabus could build several examinations, each one providing a different emphasis and different educational objectives and goals. It is for these reasons that the contacts between curriculum builders and the examiners must be continuous, and at all stages of development.

DEFINING OBJECTIVES

To establish contact between curriculum planners and examiners it is essential that the objectives and goals of education, both general and specific, should be clearly defined and clearly set out. Without clear direction as to the goals and objectives in education it can follow that there can be the situation where the main emphasis will be teaching what is tested rather than testing what is taught. Examinations can either lead or follow in education. When examinations become the key determiners of curriculum and education, it is usually by default, because the curriculum planners and syllabus writers have not been clear enough in their directions and objectives. Similarly if objectives are clearly defined but examiners are not properly informed about these and cannot translate them into the examination material the same unhappy situation may occur.

In order to meet today's needs, curriculum panels or testing panels must be conversant with:

(1) modern testing ideas;

(2) general aims and objectives of education for the country;

(3) the desirable objectives, spirit, and emphasis for that particular syllabus or subject.

CONCLUSIONS

Curriculum and examinations are two sides of the same coin and it is only when they operate together that the goals and objectives of education can be adequately reached.

The pre-requisites, then, are:

(1) that curriculum makers must build into the original curriculum evaluation and testing goals and objectives, and

(2) see to it that they reach the examiners who actually make up the tests;

(3) that the examiners (i.e. the examining body) must keep themselves fully informed at all stages of curriculum planning and defining objectives, and

(4) become conscious of the spirit, aims and implications of the written syllabus;

(5) that both curriculum planners and examiners, working as a team, appreciate where they are leading and heading from the earliest stages in terms of what will be measured and how.

PART II

REVIEW AND EVALUATION:

As indicated in Part I, the focus today in progressive education is on individualizing education, focussing on the individual. This means getting to know the pupil. Knowing the pupil in turn requires a number of practices including the necessity of measuring and evaluating each pupil in order to recognize individual differences. Hopefully the day of considering examinations as something separate and apart in education is over. The curriculum, the teaching and learning process and test and evaluation procedures should all be part and parcel of one complete and integrated process. Evaluation procedures turn education from a teacher-to-pupil 'monologue' into an effective 'dialogue' and communication. Continuous evaluation makes the process truly fruitful and meaningful.

Two main reasons have been given earlier for having examinations: the first is the need to provide evidence of an individual's competence to move from one social status to another; the second is to provide, as part of a continuous process of education, a method by which the teacher can assess what each student has and has not learned.

An analogy to educational assessment could be the study of cybernetics, a comparatively new science which is concerned with the behaviour of control systems in the physical and biological worlds. Perhaps the basic law of this science is that goal-seeking systems are error-actuated. What does this mean? Here is an example: a missile which "homes" on a target does not, in fact, go straight to it. Its course is constantly changing and its direction is modified in accordance with "feedback" information about its errors. The missile receives feedback which tells it how it is off target, and it then makes appropriate compensatory movements, though it can never be said to be exactly "on target". The same process can be recognised in

many human goal-seeking activities, and it seems directly applicable to the educational process. The teacher can apply this process consciously if he draws up a list of objectives which he hopes to achieve (he hopes to effect certain changes in his students as a result of the course he is teaching) and then by frequently evaluating his pupils' progress. He then has the feedback information which is necessary to reduce the errors inherent in progress towards any goal.

Most control systems have an optimum frequency for receiving feedback, they will swerve this way and that in their progress towards their target, but if they receive feedback too frequently, they may be unable to process the data at sufficient speed. (e.g. the feeling when one set of essays is due for collection before you have finished marking the previous set?) So the first question is: How frequently should one obtain such information? As we shall see later the answer is, as often as possible, in fact, continuously (but systematically).

It is a common occurrence to find the classroom teacher most surprised at the results of a testing programme. Assuming the test is a good and valid one, it shows how often teachers know little or nothing about the progress and capabilities of individuals, or indeed of even a whole class. Perhaps the most vital point in sound educational evaluation is the fact that to be effective evaluation must be both diverse and continuous. Part I of this paper, on the role of testing, emphasizes that tests are not precise instruments and that even the best of tests under the best conditions leaves large margins for error. For the sake of accuracy and reliability alone, evaluation must be continuous and not just a periodic event for selection or admission. However, as has been pointed out already, testing should be the other side of the educational coin, the means whereby communication and feedback to the educators is established. So, in order that evaluation procedures may provide accurate information to (1) pupil, (2) teacher, (3) parent and (4) administrator, and in order that evaluation may become an integral and useful part of the educational process, it must be continuous.

As a periodic event applied at certain times for screening and selecting pupils, testing has a limited educational function. And this is mainly an administrative function since it allows those in charge to pick people, pass and fail people, and assign a certificate or value to the person in question. It has limited educational value because it does not necessarily help the teacher to teach better, or the pupil to learn more effectively, which is, after all, the core of the educational process. Evaluation taken in the many forms discussed later, and used for diagnostic and remedial purposes on a day-to-day basis is the type of evaluation which really is a valuable part of the teaching-learning process. In this sense, written tests, oral questioning, quizzes, projects, reports, classwork, homework, etc., etc., are all considered as evaluative measures. Records are kept and results are analysed first in the classroom to help the teacher teach better and pupil to learn better, and secondly in the larger area of planning, curriculum, etc.

This is not really something dramatically new. It simply involves a consciousness on the part of teacher and other educators of the need for constant feedback and communication from the pupil. It shows a mature realization of the weaknesses and deficiencies of any one test or group of tests. It simply means a greater emphasis on gaining more and better information about the pupil and what is happening in the classroom and then using this in the next stages. It also involves a recognition of the importance of individual differences and the realization that to know pupils well and

accurately, we must assess (1) very frequently and (2) in as many ways as possible. It means that the focus is taken off the class as a single entity which must absorb a certain amount of material and be able to regurgitate this in an examination. Instead, education with evaluation as an integral part focusses on:

(1) developing the full potential of each individual to his or her capacity;

(2) seeing education as a dialogue between the teacher and the learner, where both, communicating effectively and constantly, 'grow' together;

(3) providing more accurate information as a result of continuous, diverse, and multi-faceted evaluation to all the people who need it - first of all the teacher and pupil, secondly parents, thirdly curriculum builders, planners and administrators.

But to be successful, a wide-scale use of testing and evaluation as a continuous element in the schools requires certain basics and pre-requisites. The main one is in the training of the teachers both in training institutions and through in-service work. Before testing can play the role described above, an understanding of some testing theory along with enough technical knowledge to understand, use and interpret tests and evaluation procedures must be basic to all education officers and teachers. They must be able to make effective use of objective as well as essay questions and to be able to use all the other evaluation techniques. In the field of objective tests they should have a practical classroom knowledge about the construction and interpretations of the various types of objective tests such as multiple choice, fill-in blanks, true-false, matching and so on. In addition a "guidance approach" to the child and test results and cumulative records should be a part of the teacher's equipment.

It should be emphasized as well that testing and evaluation does not mean only large scale sophisticated standardized tests. Widespread use of classroom testing combined with effective records and use of results, could provide much of the evaluation data now lacking. "Measurement devices and techniques prepared by the teacher are often the best and sometimes the only means of determining how well a class or individual pupils are progressing towards the objectives of instruction."(3)

But in such continuous classroom testing the teacher must develop a certain level of sophistication. For instance, the ability to analyse test results (do a simple item analysis) can add a great deal to the teacher's knowledge of the effectiveness of his teaching and the extent of the learning. It is also a powerful tool for test improvement. Item analysis also indicates which items are too easy or too difficult to discriminate between better and poorer examinees and it can be done simply and with little loss of time in the classroom. Then, too, teachers must understand validity in testing - that is, that the test measures what it is supposed to measure.

---

(3) Victor H. Holl, Introduction to Educational Measurement, Haughton Mifflin & Co. Boston, 1965.

The classroom teacher who has a sound knowledge of testing and who has available cumulative test records is in a far better position to understand the learning problems and difficulties of individual children. He is able to identify the most capable youngsters, who need enriched learning experiences, as well as the slow learners who may need special help and modified assignments. The slow learner who achieves less because he is slow mentally is a perfectly normal child; he should in no sense be considered a failure simply because he does not reach the average level of achievement of children of his own age or grade. On the other hand, the child with high ability who does mediocre work, is, in a truer sense of the word, a school learning problem. The concept of failure in school is one with which we could very easily dispense since it is never possible to determine with certainty who is failing - it may be the school quite as much as the child.

It should be stated emphatically that standardized testing here is no complete substitute for an effective evaluation programme on the part of the classroom teacher. Such an evaluation programme includes the teacher's own locally constructed tests as well as ratings on specially assigned projects and daily classroom recitations. Nevertheless, the professional technicians who develop standardized tests can offer the classroom teacher many suggestions for evaluating the results of classroom instruction. Indeed, much in-service training is needed in this important area.

At several points it has been mentioned that evaluation must (1) be continuous (2) employ a wide variety of evaluative procedures. What are some of these ways of evaluating? Here is a list of some of the many common-sense methods for diverse and multi-faceted evaluation:

(1) Tests - there are many types including:

  (a) Achievement:

    (i) informal teacher-made

    (ii) standardized

  (b) Mental ability

  (c) Personality

  (d) Aptitude

  (e) Interest;

(2) Rating scales;

(3) Checklists, surveys, inventories and questionnaires;

(4) Observation;

(5) Records and reports:

  (a) cumulative folders,

  (b) anecdotal reports,

  (c) diaries and logs;

(6) Interview;

(7) Sociometry;

(8) Role-playing:

(a) sociodrama,

(b) psychodrama;

(9) Situational or performance tests;

(10) Student papers and projects:

(a) papers,

(b) notebooks,

(c) reports,

(d) autobiographies,

(e) personal data sheets;

(11) Case studies;

(12) Case conferences.

Tests must measure all the important outcomes of instruction such as course objectives, factual knowledge, understanding of human nature, the proper weight for each topic and so on. Benjamin Bloom[4] in his taxonomy lists six main objectives or outcomes of learning in the cognitive domain that should be measured. These include:

(a) Knowledge of specifics, ways and means of dealing with specific universals, abstraction from specifics;

(b) Comprehension, involving abstraction, interpretation, extrapolation of communication;

(c) Application of knowledge;

(d) Analysis of elements, relative principles;

(e) Synthesis;

(f) Evaluation.

Before continuous evaluation becomes a full and integral element in education there must be clearly defined educational objectives. Good evaluation shows how far school progress meets the objectives set out. These must be clear both to the policy makers, examination bodies and teachers. Unless these are clarified, testing cannot play its proper role.

---

(4) Bloom, B.S. (ed) - Taxonomy of Educational Objectives, Handbook I: Cognitive Domain, David McKay & Co., Inc., New York, 1956.

Then too in newly developing countries there are other problems which often require close attention before testing can be effective; these include:

(a) Keeping records. Records of pupils' ages are basic to many kinds of testing. Continuous evaluation is only useful when a cumulative record of each individual is kept and used.

(b) Both teachers and pupils must become familiar and at ease with such things as objective tests, the concept of "time tests" where every minute counts, and so on.

(c) Often in such countries there is a less settled population of teachers and pupils, both of whom move about and leave schools frequently. This makes for difficulties in establishing norms, control groups and experimental groups.

In newly developing countries differences between urban and rural groups tend to be bigger. Urban groups quickly become sophisticated in a variety of aspects and ideas while rural pupils remain almost totally unaffected. Language problems, administrative problems and others must all be tackled in a special way for rural areas.

It was mentioned earlier that in addition to evaluation being continuous it must include a wide variety of techniques and methods. In testing, we are measuring people and their responses and knowledge and not bricks or bridges. Physical things can be measured accurately and completely with a ruler or scale. Because of the complexity of man and the complexity of the facets we wish to measure, there is more chance of accurate assessment if a variety of techniques are used and used often. And by accumulating and combining results, we are more likely to measure accurately the many processes and facets that we see as the goals of education.

In summary, then, worthwhile testing should meet a number of criteria and requirements. An attempt is made here to include the key elements in a graphic form:

GOOD WORTHWHILE TESTING

Proper Administration

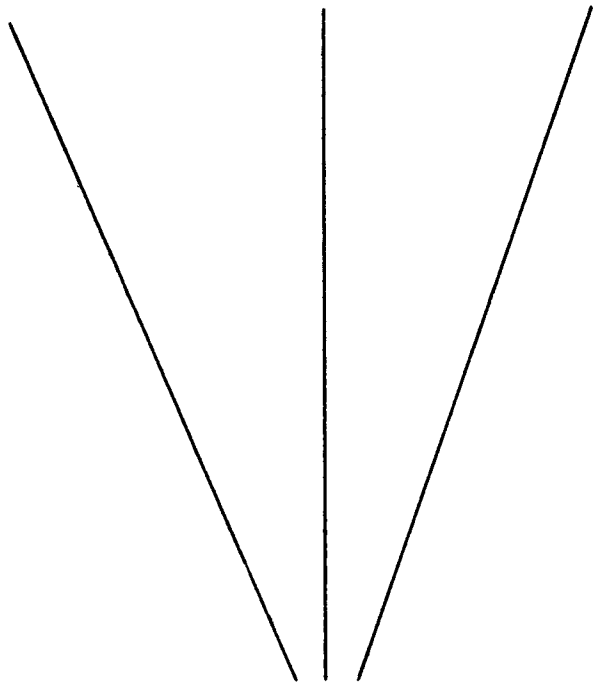
Timing, testing conditions, uniformity, etc., are all properly done.

Good Sampling, Clear Objectives

Questions measure different aspects and levels (Bloom) - valid and reliable tests - the tests measure all the objectives of education and measure it always in the same way.

Continuous Evaluation

A variety of types of ways of testing are used continually as an integral part of the teaching and learning process. Testing must be continuous:
(1) to provide communication and dialogue between teacher and pupil;
(2) to make evaluation an effective teaching and learning tool by providing pupils and educators with feedback information on what is happening;
(3) to allow pupil and teacher to assess progress, assess their work and make adjustments;
(4) to provide up-to-date information to teachers, pupils, parents, administrators, curriculum builders and policy makers;
(5) to compensate for the inherent weaknesses in the results of any one test. The average results of a great many (continuous) assignments are much more valid, reliable, accurate and meaningful than the results of any one or two major tests however carefully constructed.



PROPER USE OF TESTS

Interpretation

Test results are interpreted for what they are. No more or less evidence or value is attached to them than they deserve. Test results are not regarded as "god-like" nor as "useless". They are given their proper due and right and used as one piece of information along with all other information available.

Well used information

A truly useful test provides information to all of:
1) pupils
2) teachers
3) administrators and inspectors
4) parents
5) school curriculum builders and policy makers.

Variety of Technique

Evaluation must take a variety of forms and include as many different kinds of evaluative techniques as possible. These should range from carefully constructed standardized terms to common-sense evaluative observations and ratings of teachers. This is because
1) we are assessing complex human beings;
2) the qualities being measured are tangible, abstract and difficult to measure;
3) more kinds of measurements taken more often ensure more accuracy, reliability and validity.

Part I of this paper presented some points to stimulate discussion on the topic of the nature, place and influence of tests and measurement procedures including examinations; Part II has focussed on continuous review and evaluation. As was pointed out in the opening statements of Part I, the author feels that the two topics are indeed only one. Evaluation is a continuous and integral part of education. The frequent repetition in Part II of points from Part I are meant to emphasize this point.

However, for organizational purposes, the paper was divided into two parts. It is hoped, however, that the overlapping and repetitions of similar points in the two parts may lead to a line of thinking that combines the two ideas.

This marks the end then of a few brief ideas in the field of measurement. Not all points are covered, nor are those that are mentioned covered adequately or completely. However, hopefully, enough has been said to provide the raw material so that discussion can whittle away the rough edges and produce a refined and finished product on this important and contraversial subject. Perhaps because of its deficiencies, this paper will serve the better to stimulate discussion, which is, after all, its purpose.

TESTING WITH EDUCATIONALLY DISADVANTAGED CHILDREN

A.E.G. Pilliner,
Director, Godfrey Thomson Unit for Academic Assessment,
University of Edinburgh, Scotland.

Every country has its educationally disadvantaged children, even those in which educational development is most advanced. Britain is now replacing a selective system of education by a comprehensive one in an endeavour to eliminate, or at least reduce, unequal educational opportunity. Superimposed on this task, she is now faced with the responsibility of educating an increasing number of young immigrants from other Commonwealth countries. The United States of America, despite the fact that in principle her educational system has never been other than comprehensive, has not yet achieved her avowed aim of de-segregation and the quality of education offered to some of her citizens is still inferior to that enjoyed by others.

It should not surprise us that this state of affairs exists a fortiori in countries at an earlier stage of development. In such countries, through sheer force of circumstances, education in any way comparable in quality and adequacy to that taken for granted for the majority in some advanced countries is available to only a small minority. The allotment of a large proportion of scarce resources to the more extensive education of a relatively small proportion of children is understandable. Pre-requisite to speedier progress in the future, further technical advance, increased economic development and wider educational expansion, is the production now of a necessarily small number of people possessing the knowledge, skills and dedication essential to the achievement of these aims.

It is no accident that in many of the new countries educational objectives tend increasingly to resemble those in others more fortunate in having advanced further along the path of development. According to Doob (6), the pressures forcing the new countries in the same direction are inevitable, irresistible and irreversible. This does not mean that all will arrive at the same place. A country on its way 'up' will be selective in what it absorbs and will adapt its acquisitions from elsewhere to its own traditions and needs. Nevertheless, since both less and more developed countries share a number of the same objectives, their educative processes will have much in common. At the same time, a process evolved over a lengthy period and geared to the norms of a society or culture already well developed cannot be transferred ready-made to another less so without giving rise to problems, even though the objectives are similar.

These problems are reflected in the testing procedures which are an integral part of any and every educational process. The situation previously mentioned implies that a relatively small number of pupils must be selected for secondary education from a very large primary school population. Countries where this situation exists are likely to be characterised by primary education of poor and uneven quality. Children living in towns may be more fortunate in their primary education than others living in villages.

In this case, restrictions are imposed on the interpretation of scholastic attainment test results. Though such tests may still accurately

measure a pupil's achievement in specific subjects to date, their use as prognosticators of future success is precluded or at least limited, however successful they may be in this respect with children more fortunately circumstanced. The poor performance on an arithmetic test of a pupil who has hitherto been taught arithmetic either badly or not at all is a fair index of his present ability in that subject. But as a predictor of his likely progress if this defect is remedied its value is questionable. If assessment of potential or aptitude is at issue some other means must be found.

If experience in countries more educationally advanced is anything to go by, the use of tests of verbal reasoning might seem to offer a solution. Tests in this category differ from tests of scholastic attainment in that they are less closely geared to the school curriculum; good performance on them is less dependent on exposure to the usual range of school subjects. They have been extensively used for 11+ selection in Britain, where numerous follow-up studies have consistently shown them to be among the best predictors of academic success.

However, difficulties still remain. There may be several native languages or dialects while the accepted medium of instruction in the secondary schools is a second language such as English, the pupil's acquaintance with which is limited by factors such as his primary teachers' command of it.

Bernstein's (3) work on language habits in Britain bears on this situation. He points to the relation between class structure and the varieties of English used by school children. Social stratification is related to differential availability of language codes. The lower working class child has a group-oriented 'restricted' code; the middle class child has both this and an individually-oriented 'elaborated' code. These codes differ in that the first is more fluent, repetitive and predictable, the second more hesitant, idiosyncratically planned and complex. Educationally, the child from a poor background is at a disadvantage since he finds himself having in effect to translate what he hears from his teachers. As Bernstein points out, differential difficulty in communication is likely to be reflected in differential verbal test performance.

The problem is exacerbated when the differential is not merely intra- but inter-language. It is therefore natural to consider the possibility of assessing pupils' aptitude for further stages of education by some testing procedure which avoids the use of the differentially unfamiliar second language. On the face of it, one way of doing so would be to couch the tests employed in the pupils' own native languages. This however may be difficult in practice if several languages are involved. There is the further technical difficulty of equating the performances of different children on different tests - for, let there be no mistake about it, even the same content translated into different languages produces different tests, the results of which, expressed numerically, are not necessarily comparable. Moreover, the problem of unequal primary school opportunity, and its implications for scholastic attainment, will still remain.

On all these counts, it may be thought desirable to go one step further, to eliminate the use of language so far as is practically possible, and to rely on non-verbal or non-language tests. Here, surely, it might be argued, is the way out of the difficulty. If the use of language-bound tests is seen as impracticable or leading to injustice, should not their substitution by non-language tests reduce the practical problems and promote

'fairness' for all concerned?

This is the kind of thinking behind the more general concepts of 'culture-free' and 'culture-fair' testing. The intention is wholly admirable. Any measure which will help to redress the balance in favour of children who are culturally deprived or otherwise educationally disadvantaged is surely to be encouraged. The laudable objective is to reduce these obstacles by the use of testing devices which transcend or remove cultural differences or educational inequalities.

However, the problem is by no means simple. The concept of 'culture-free' tests is highly dubious. Anastasi (1, p.256) is surely right when she says: 'No test can be truly "culture-free". Since every test measures a sample of behaviour, it will reflect factors that influence behaviour. Persons do not react in a cultural vacuum.' Wesmen (16, p.269) is even more forthright. 'I do not wish to impugn the high social motives which stimulate the search for such devices; I do wish to question that such a search, in its usual setting, is sensible. A culture-free test would presumably probe learnings which had not been affected by environment; this is sheer nonsense.' These statements represent the general view of most contemporary psychologists. Few would now regard the quest for culture-free tests as other than chimerical.

The prospect for 'culture-fair' tests is, on the face of it, less unpromising. In principle it is possible to build tests which, though not free of cultural influences, sample only behaviour common to several cultures. An alternative description of such tests is 'cross-cultural'. The amount of effort that has gone into the construction of allegedly cross-cultural tests is vast, particularly if we include also tests intended for comparisons among sub-cultures within a larger culture. Only a few can be mentioned here. In the nature of things, they are non-verbal in content. They fall into two main categories: performance tests, designed for individual administration, and in the main involving manipulation of objects; and non-verbal or non-language group tests, normally paper-and-pencil tests which do not demand of the testees the skills of reading and writing. Most such tests do however depend on oral instructions, it being assumed (perhaps too lightly) that these are of such simplicity that no semantic problems arise in their translation and that different language versions do not differ in difficulty. A few tests have been constructed in which the instructions can be mimed or demonstrated.

Examples of tests in the performance category are: Form-board (Sequin, Pintner-Paterson), Mazes (Porteous), Picture Completion (Healy), Block Manipulation (Kohs); Stencil Design (Arthur); Analogies (Leiter); and, of course, the General Performance Scale of the WISC (Wechsler). Examples from the group non-language category are the Draw-a-Man (Goodenough), Matrices (Raven), Pictorial Problems (Davies-Eells), Semantic Symbols (Rulon); and a number of tests intended to probe, using pictorial or diagrammatic material, mental functions - analogies, odd-man-out, series and the like - similar to those frequently occuring in verbal tests (Moray House Picture, Jenkins Non-Verbal, Cattell IPAT).

On closer examination, however, the prospect of producing 'culture-fair' tests is only slightly less unpromising than for tests that are 'culture-free'. By restricting test content to elements common to several cultures the relevance of the results in respect of any one of them is made questionable. To the extent that different cultures display unique features,

nurture disparate traditions and values, or foster or suppress different abilities or modes of behaviour, tests restricted in this way may miss their targets. To quote Anastasi (2, p.299) again: 'If we were to rule out cultural differentials from a test, we might thereby lower its validity against the criterion we are trying to predict'. It is as though in trying to please everybody, we succeed in pleasing nobody. Or, to change the metaphor, although the wave pattern for the fundamental tone emitted by different musical instruments is the same for all, it is the superimposed over-tones or harmonics which endow each with its peculiar timbre, its richness of quality.

The concepts of 'culture-free' and 'culture-fair' tests once received plausible support from the contemporary psychological theory. 'Native intelligence', like original sin, was reified and came to be regarded as a fixed entity rather than a developing attribute. By the exercise of sufficient inventiveness - Wesman (16) speaks of 'ingenious mining devices' - the influence of differential exposure to learning could be eliminated and the 'innate intelligence' of the individual revealed and recorded on a scale for all to see.

More recent theory is less accommodating. Hebb's (9) distinction between Intelligence A and B corresponds broadly to the geneticist's distinction between genotype and phenotype. Like the genotype, Intelligence A is not directly observable, still less measurable. Only Intelligence B, corresponding to the phenotype, can be observed; it results from the interaction of both nature and nurture. The title of a once popular song sums it up neatly:; 'It's what you do with what you've got that counts'. Vernon (14) playfully, in the first place, one suspects, but then more seriously, had added a further category. Intelligence C is what tests measure. It varies with difference in test content and is therefore not unique in the prediction it affords of Intelligence B. Hebb's theory offers but cold comfort in the search for instruments equally fair to differentially disadvantaged testees.

On the fact of it at least, the theory of 'fluid' and 'crystallised' intelligence attributable to Cattell and Horn (5) is distinctly more hopeful. They suggest that the general factor emerging from studies of batteries of disparate tests is a mixture separable into two components: G_f ('fluid' intelligence), reflecting constitutional equipment; and G_c ('crystallised' intelligence), the results of experience such as cultural and educational pressures. Unlike Intelligence A, G_f is measurable by tests tapping adaptability to situations so unfamiliar that previous learning experience is of no help. G_c, corresponding roughly to Intelligence B, is manifested in cognitive behaviour already patterned by previous experience. Even before biological maturity is reached, diversity in cultural opportunities, interests and personality traits produces substantial individual differences in G_c which, according to the theory, should not be parallelled for G_f.

This theory underlies the construction of the Cattell IPAT Culture Fair (formerly Culture Free) Intelligence Test. Predictably, the greatest success in removing 'contamination' by cultural differences is claimed for subtests involving mazes, identification of similar drawings, picture classification and symbol copying. At best, however, the success achieved is only partial. In view of the IPAT, Tannenbaum (12, p.454) concludes that 'the goal of demonstrating equality among national and international subpopulations by some measures of general ability has not been reached

by this test.' He questions whether this is a goal worth pursuing. 'Even if it were possible to devise a test so antiseptic as to clean out inequality not only among subcultures but also among other groups showing differences in test intelligence, such as those classified by sex, age, geographic origin, body type, physical health, personality structure, and family unity - what kind of instrument would we have then? Since such a test must perforce be so thoroughly doctored as to omit tasks that reveal these group differences, or substitute others that show "no difference", what could it possibly measure? What could it predict?' Vernon's (15, p.25) conclusions are equally definite. 'The main weakness in his (Cattell's) theory is the claim that fluid ability tests are largely immune to cultural influences. The skills required for reasoning with these abstract materials would appear to be built up in just the same way as those involved in verbal reasoning; and the evidence ... demonstrates at least as great variation attributable to cultural differences'.

For a very complete and up-to-date survey of this evidence, reference should be made to Vernon (15). Only some of it can be cited here. As already stated, the IPAT was found to be only partially successful in ironing out cultural differences. Although in cultures similar to that in which the test was developed the same norms were approximately applicable, this was not so for cultures more dissimilar; for these, average performance was often much lower. Bernstein (4) reports smaller differences in performance on Raven's Matrices between middle and working class groups than on tests of verbal reasoning. But in other studies, particularly in African countries, test results were positively correlated with amount of education. The Goodenough Draw-a-Man (8) test has gone through several revisions. After extensive use with a number of different cultural and ethnic groups, its authors have abandoned their original optimistic view and in their more recent reports have concluded that a culture-fair test of whatever attribute 'is illusory'.

The Davis-Eeels Games (7) were specially designed for American use to be relatively independent of social class bias. But differential educational disadvantage was still reflected in differential performance on these tests no less than on more conventional intelligence tests which were in addition more predictively valid in respect of tested achievement and teachers' assessments.

One of the most interesting and definitive studies in this area is that conducted by Ortar (10). She administered both a Hebrew version of the WISC Verbal Scale and also the Performance Scale to upwards of 1000 Israeli children. These were divided into five groups with different cultural backgrounds ranging from recently arrived Oriental immigrants to an Israel-born 'high status' group (mainly of European parentage). After re-standardising both Scales for Israeli children, she found the 'cultural distances' between the groups to be larger on the Performance than on the Verbal Scale. In a similar study conducted with Scottish children Tsakalos (13) found differences in social status to be reflected in differential performance on the Jenkins non-verbal test no less than on Moray House tests of verbal reasoning and scholastic attainment.

The conclusion is inescapable that it is fruitless to search for testing instruments that will somehow transcend cultural differences and educational inequalities. What are the implications?

In the first place, it must be recognised that belief in the essential

equality of man receives little support from the considerable research in this area which it has stimulated. It remains an act of faith. This need not deter us from acting on that belief. A warrant from psychologists qua psychologists is not essential to the maintainance of a fundamental principle on which the advance of civilisation is predicted.

Secondly, it has to be accepted that educational disadvantage is endemic and that there is no simple counter to it by way of tests purporting to reveal intelligence, talent, potential, or whatever we may choose to call it, irrespective of differences in cultural, social or educational background. Such tests are of dubious value to a primary school teacher in Britain faced with an influx into her class of immigrant children without a word of English among them. There is no simple way of helping her to differentiate among them, or between them and their native-born peers, in terms of 'basic' intelligence. Her best practical policy still is to do all she can to make them feel welcome and to teach them English. Likewise, such tests offer no panacea to a developing country where, because of scarce resources, stringent selection is necessary and too many children are chasing too few places in the educational sun. The brutal truth must be faced that there are plenty of other children whose claim for preferment is no worse than that of the fortunate few selected. The solution to the problem is economic, not psychometric.

From an educational stand-point, the best hope of advance in general, and amelioration of educational disadvantage in particular, lies in the field of language-teaching. The mother-tongue may suffice if it provides for effective communication with other nationals and is suitable as a medium for advanced education. If not, a second language is necessary, taught, as Vernon points out, not peripherally, but as a central tool of comprehension and thought.

What then should be the role of the psychologist? There is no reason why it should change materially, though possibly a shift of emphasis is indicated. Any still engaged in the search for testing instruments equally 'fair' in different cultures should bear in mind the fruitless quest of the alchemists for the philosopher's stone; though they may console themselves by reflecting that (in a different sense from the original alchemists') the transmutation of metals has now been accomplished. There is a lesson here. That achievement was the outcome of 'pure' research not specifically aimed at transmutation, nor concerned with its consequences. So too with the psychologist. He should listen to Anastasi's (2, p.302) warning: 'It is not (the psychologists') role to provide ready-made solutions to insoluble problems. It might be salutory if testing gave less heed to the pull of practical needs and more to the thrust of behavioural sciences'.

But less heed is not the same as no heed at all. The psychologist, like the physicist, has responsibilities outside his laboratory. Despite all that has been said, he has yet much to give in the field of testing in the service of education. It is a truism that the best indicator of a child's learning potential is a test sampling previous learnings which are relevant to the criterion or criteria we wish to predict. For long enough this maxim has guided with reasonable success the construction of tests for educational purposes within western cultures. There is still room for further research of the kind that Schwarz (11) has engaged in, aimed at discovering the previous relevant learnings in cultures elsewhere in a stage of transition.

Let Vernon (15, p.229) have the last word. 'What is important is that in concentrating on abilities recognised by western cultures,

psychelogists should not neglect special talents that might be more highly developed in other countries'. To extend a metaphor employed earlier, in seeking out these special talents we may be taking a small but useful step towards the assembly of a cross-cultural orchestra.

REFERENCES

1. Anastasi, A. (1964) — Psychological Testing. (2nd Edition). New York: Macmillan.

2. Anastasi, A. (1967) — 'Psychology, Psychologists, and Psychological Testing'. In American Psychologist, 22,4, pp. 297-306.

3. Bernstein, B.B. (1961) — 'Social Class and Linguistic Development: A Theory of Social Learning'. In Education, Economy and Society (Halsey, A.H.) Glencoe: The Free Press, pp. 288-314.

4. Bernstein, B.B., and Young D. (1966) — 'Some Aspects of the Relationships between Communication and Performance in Tests'. In Genetic and Environmental Factors in Human Ability (Meads, J.E. and Parkes, A.S.). Edinburgh: Oliver and Boyd, pp. 15-23.

5. Cattell, R.B. (1963) — 'Theory of Fluid and Crystallised Intelligence: A Critical Experiment'. J. Educ. Psychol. 54 pp. 1-22.

6. Doob, L.W. (1960) — Becoming More Civilised. New Haven: Yale University Press.

7. Eeels, K., Davies, A., Havighurst, R.J., Herrick, V.E., and Tyler, R.W., (1951) — Intelligence and Cultural Differences. Chicago: University of Chicago Press.

8. Goodenough, F.L. and Harris D.B. (1950) — Studies in the Psychology of Children's Drawings': II. 1928-1948. Psychol. Bull. 47, pp. 369-433.

9. Hebb, D.O. (1949) — The Organisation of Behaviour. New York: Wiley.

10. Ortar, G. (1963) — 'Is a Verbal Test Cross-Cultural?' In Scripta Hierosolymitana. Jerusalem: Magnes Press, The Hebrew University, pp. 219-35.

11. Schwarz, P.A. (1961) — Aptitude Tests for Use in Developing Nations. Pittsburgh: American Institute for Research.

12. Tannenbaum, A.J. (1965) — Review of IPAT Culture Fair Intelligence Test. In Mental Measurements Year Book (Ed. Buros, O.K.) New Jersey: The Gryphon Press, pp. 453-4.

13. Tsakalos, P. (1966) 'Is a Non-Verbal a Culture Fair Test?' Master of Education Thesis (unpublished). University of Edinburgh.

14. Vernon, P.E. (1960) Intelligence and Attainment Tests. London: University of London Press.

15. Vernon, P.E. (1969) Intelligence and Cultural Environment. London: Methuen.

16. Wesman, A.G. (1968) 'Intelligent Testing'. In American Psychologist, 23, 4, pp. 267-274.

INFORMAL TESTS FOR CLASSROOM USE

W.B. Elley,
Assistant Director
New Zealand Council for Educational Research.

WHY TEST?

In the course of his career, every teacher will have occasion to prepare, administer and mark hundreds of tests of his pupils' attainment. Sometimes these tests will consist of carefully selected formal written exercises with a rigid time limit and an elaborate marking system; sometimes they will be spontaneously constructed, orally presented and evaluated only superficially. But all will have these characteristics which distinguish them from external examinations and standardized tests:

(i) Classroom tests are prepared by the teacher or headmaster for local rather than national use. They are usually prepared at short notice, without the benefit of special expertise, panel review, or pretesting of questions.

(ii) Classroom tests are usually designed to evaluate the pupils' mastery of short units of work recently studied, or objectives which are specific to a school, a class, or a lesson. A teacher may prepare a short test on multiplication and division with decimals, or the causes of World War II, or a list of irregular French verbs. By contrast, an external examination usually evaluates the fruits of a year's study - or more, while a standardized test normally samples basic objectives which are developed over an extended period of time, and are not dependent on the teaching of a particular course.

However sophisticated his test preparation procedures, it should be apparent to all that a teacher requires the results of classroom tests to make decisions about his teaching:- whether to proceed or to back track, whether to change a teaching method, or to introduce a new topic. Judgements have to be made about classification and selection of pupils, advice must be offered about course changes and vocational plans, suitable materials and approaches must be found for children at all levels of ability, pupils must be identified for special treatment. Whether a teacher's decisions are required for groups or for particular individuals, they are more likely to be sound if they are based on accurate information about the abilities and attainments of his pupils. If this information is to be helpful, it should be obtained from tests which are both reliable and valid. Tests which are too easy or difficult, tests which are too short or too long, tests which sample only part of the course, or which weight certain parts too heavily, tests which are ambiguous in their directions, or which leave too much to chance, tests which cannot be marked with reasonable objectivity - such tests may mislead both teacher and pupil, confirm erroneously-held prejudices, and occasionally lead to injustices with far-reaching effects. How can teachers prepare classroom tests which will produce results in which they can place confidence? What test construction methods are likely to produce tests of adequate reliability and validity? First we must examine these criteria of a good test. What does it mean to say that a test is reliable and valid?

RELIABILITY.

Tests are reliable if they produce consistent results, if they produce similar marks on different occasions. If a pupil gains 100% in a foreign language dictation test today, and only 50% tomorrow, then the results are not consistent, the tests are not sufficiently reliable to base judgements on. If a pupil is placed first in his class in a test of multiplication and division of decimals on one occasion and is 20th in a subsequent test of the same skills, we can conclude that the tests are not reliable indicators of his ability.

To be reliable a test must normally be long enough to minimize the effects of chance factors in the content and skills included in the test. With a short test, a pupil may be lucky, because he happened to know or guess correctly the few questions that were asked, whereas he knew very little about the areas untouched by the test. A standardized test of reading, mathematics or language, normally requires at least 40 sound objective-marked questions to reach a satisfactory level of reliability. To make decisions about individual pupils, a teacher-made test will probably require more questions than this. For judgements about groups, a teacher may get by with fewer. Just how long a particular test should be depends on the type of material tested, the amount of supplementary information available, and the importance of the decisions being made. Thus a test of a highly specific skill, such as arithmetical computation, or typing, may produce reliable results within ten minutes. If however, we wish to examine a pupil's grasp of mathematical relationships, or his understanding of a period of history and to make decisions about future schooling on the basis of the results, we may wish to extend the test over two hours to gain maximum reliability. For such skills as essay-writing ability, or oral expression, it is commonly found that pupils vary so much in their performance from day to day that the only way to gain adequate reliability is to test the pupils on several topics (over several occasions), and to combine the marks given by two or three independent markers.

Other requirements of a reliable test are clear, precise directions and reasonable time limits. The questions should be unambiguous, neither too easy nor too difficult; they should discriminate well between good and poor pupils, and they should be capable of reasonably objective scoring.

VALIDITY

A good test must be valid. This means that, in addition to measuring a pupil's attainments reliably, it should be relevant to the needs of the tester. It should cover the unit or course adequately, sampling each content area and skill in appropriate proportions. If a teacher knows precisely what his objectives are, he can usually tell, by analysing the questions of a test, whether they conform closely to the objectives he has adopted i.e. whether the test is valid for his purposes.

To illustrate, a 100-item test of mathematical computation may be highly reliable, and yet be quite invalid for measuring achievement in a course of modern mathematics which emphasizes concepts, relationships and reasoning. The objectives of the test do not match the teaching objectives. Again, a test of geography which focusses on isolated details about populations, areas, climate, exports, capital cities and the like, would produce irrelevant results for a teacher who stressed broad concepts, generalized skills and underlying relationships. A valid test of such

objectives may require novel or fictitious situations on which to base questions so that a pupil can demonstrate that he has attained these objectives, regardless of the particular factual details he has acquired.

To ensure maximum validity for his tests, then, it is important for a teacher to spell out, as clearly as possible, precisely what his objectives are, and to build his questions around these, in the appropriate proportions. Tests which develop without such planning often degenerate into factual quizzes of the low-level, isolated, easily testable fragments of the course.

DEVELOPING THE TEST.

(i) Once a teacher has decided on the purpose for his test, he should consider the various objectives he has in mind, and how he might best classify them. For a content-oriented course, such as science or history, he might first divide the course into the main content areas, and ensure that each receives a fair ratio of questions. A general science course may be classified into three main areas - say chemistry, physics and biology. A more specific classification, for a biology course, might be living organisms, life processes, conservation, heredity and reproduction, and evolution. In addition a teacher should ensure that questions test different levels of understanding. Some tests concentrate on examining for recall of specific information, some for understanding of important ideas, some for application to new situations, and so on. In a science test, a useful classification system for the objectives might be
    (a) Knowledge of facts and conventions
    (b) Understanding of concepts and principles
    (c) Ability to apply the scientific method to new problems
    (d) Knowledge of industrial applications.
In language subjects the content areas are less easily defined and it may be more appropriate to classify the objectives of the course according to the skills to be tested - reading, writing, translation, dictation, etc.

(ii) Once objectives are classified, a blueprint or table of specifications can be drawn up which sets out the content areas and the objectives, and allows them to be weighted on some rational basis, before the test questions are prepared. For maximum validity, a test will normally weight most heavily those topics or objectives which have been given most emphasis in the course or unit taught. But all areas should be tested where possible.

An example of a fictitious test blueprint is set out below.

Sample Blueprint for a Mathematics Examination

Objectives	Nos.& Numerals	Measurement	Fractions	Geometry	Total
Knowledge of terms, facts	10	5	5	5	25
Understanding of concepts	10	5	5	10	30
Routine calculations ..	5	5	5	0	15
Application to new problems ..	15	5	5	5	30
Total emphasis ..	40	20	20	20	100

(iii) The third stage in developing the test requires a decision on the form of the questions to be asked. There is no question type ideally suited for all purposes. For instance, short-answer questions which require pupils to fill in the blank or complete a sentence are useful for covering a wide range of facts in a short time. Outside of mathematical subjects they are less useful for estimating depth of understanding without introducing some ambiguity in the question or subjectivity in the marking. Multiple-choice questions are widely used in standardized tests and external examinations because they can sample the whole course widely and efficiently, and test higher objectives, but such questions are not easy for classroom teachers to prepare and they do not examine the ability of the pupil to generate and organize his own ideas. Matching questions are best suited to measuring knowledge of homogeneous sets of facts or conventions. Pupils may be asked to match books with their authors, chemical compounds with formulae, countries with exports, etc. They should not be used however, unless the contents of each list form a homogeneous group, so that each item on one list is a plausible match for each item in the other list. True-false questions may have some value in classroom tests since they enable the teacher to sample widely in a short time, but they are frequently superficial, they are unsuitable if the truth of each statement is not absolute, and they are prone to be unreliable due to guessing on the part of the pupils. They can be modified of course, by requiring pupils to correct false statements, or to classify a statement as "sometimes true", depending on other factors. Perhaps their greatest value is as a starting point for classroom discussion. Essay questions compensate for some of the deficiencies of other question types in that they do require the pupil to express his own ideas, and to demonstrate fluency and organization. However, they cannot measure as many aspects of a course as do short-answer questions, and they are

difficult to mark reliably.

Before deciding on the kinds of questions to use then, a teacher should consider the various pros and cons outlined above, in relation to his own expertise in item-writing, the number of pupils involved, the time available for setting and marking, the degree of reliability required and the kinds of decisions to be made with the results.

(iv) Preparing the questions to fit the blueprint is the fourth stage. Here there are many pitfalls, and no short-cuts to success. So often when questions are hurriedly prepared they turn out to be ambiguous, too easy, too difficult, or unsuitable for some other reason. The following checklist may alert teachers to the kinds of weaknesses likely to be found in their questions.

(a) - <u>General</u>

- Keep questions brief, simple and free from complex verbal instructions, double negatives etc.

- Test only important facts and skills; avoid trivia, catch questions, and irrelevant material.

(b) <u>Completion Questions</u>

- Use a single blank in each question.

- Place the blanks near the end of the sentence.

- Ensure that there are a finite number of correct answers.

- Make all blanks approximately the same length.

(c) <u>Multiple-Choice Questions:</u>

- Use only plausible distractors.

- Ensure that there is only one acceptable answer.

- Avoid the stereotyped language of textbooks in the correct answer.

- Beware of grammatical clues and verbal associations which help the uninformed.

- Make the correct option the same length as the distracting option.

- Avoid overlap in the options.

- Avoid any discernible pattern in the correct answers.

(d) Matching Questions:

- Clarify the instructions so that pupils know the basis for matching.
- Use only homogeneous sets of items in each list.
- Make an unequal number of items in each list.
- Use fewer than ten items in each list.

(e) Essay Questions:

- Ensure that the question as it is asked cannot be more adequately measured by another approach.
- Structure the question in such a way that pupils know what to include, what to omit, and how much to write.
- Ask several short questions of different types rather than one long question.
- Avoid giving pupils a choice of questions unless it is absolutely necessary.
- Prepare a model answer before the test, but be prepared to revise it in the light of pupils' answers.

These principles may not always be applicable or even justifiable, but they do point to frequent sources of weaknesses in classroom tests. Such weaknesses can often be overcome by working with a colleague or panel of teachers, by requiring somebody to answer the questions while the test is being prepared, or by pre-testing the questions on a sample of pupils similar in ability to those for whom the test is designed.

There are no perfect paper-and-pencil tests. All are somewhat artificial; all are subject to pupil fluctuation in concentration; all provide only a sample of pupil knowledge; all therefore are to some extent unreliable. A close observance of the principles outlined above, however, should help teachers to polish up their testing procedures. Further improvements can be effected by studying textbooks on the subject, by examining well-constructed standardized tests, by item analyzing one's own tests, and by discussing one's efforts with other teachers.

EXAMINING SPOKEN ENGLISH

P.J. Hitchman,
Department of Education,
University of Nottingham, England.

For centuries the teaching of English in the United Kingdom has been very largely confined to teaching children to write their language and to study its literature from the printed text. Learning to speak their own language has been largely left to chance - the chance of social background. A middle-class house has provided much richer linguistic opportunities than a working-class environment. This has tended to a rigid stratification of class structure, with educational opportunity and career achievement unfairly tilted in favour of the English middle-class.

Now there is a belated recognition of the importance of spoken English in the world of today; and for the first time in our educational history its backing is everywhere being taken seriously.

To test a subject in our schools is to give it importance and to fix its status. Written examinations are part of our educational tradition. Now, in the last few years, we have the novelty of tests in spoken English for our school children. This is giving status to the oral language in our schools.

Official tests are in operation at three levels: General Certificate of Education Advanced Level (within the area of the Joint Matriculation Board), Ordinary Level (organised by the London University School Examinations Department), and at the level of the Certificate of Secondary Education. (C.S.E. examinations are taken in most schools in England that do not take G.C.E. O Level examinations). The J.M.B. 'A' level test in Spoken English can be taken only by those candidates taking General Studies, the London 'O' level test by those taking the English Language paper; in neither case is it compulsory. Candidates taking C.S.E. English must take a test in oral English unless excused by reason of speech defect.*

The aims of these various oral examinations can be summed up in the words "communicate, communication". This is implicit in the syllabuses of the 'A' and 'O' level authorities and is made explicit in those of most of the 13 C.S.E. boards. Probably all intend "Communication" to be understood as two-way. The North Regional Board (covering schools in the northeast of England) states: "The English Language examination will attempt to ascertain the candidate's ability (i) to communicate clearly with other people and (ii) to understand other people when they attempt to communicate with him, both orally and in writing".

The most important use of speech deliberately made audible is as a means of communication between human beings. The aims of the various

---

* In 1969 approximately 2,000 candidates took the 'A' level test in spoken English, and approximately 20,000 the 'O' level test. It is not possible to give the total for the C.S.E. test in spoken English, but in one region alone - the Metropolitan - the figure was 11,000.

examinations clearly recognise this; and their forms of examination are coming to be - as they should - a reflection of these aims.

At first the examining boards played safe. They had to their hand a well-tried test-item in reading aloud. The reading by pupil to class has been a popular form of educational activity in schools from ancient times, and teachers have always felt impelled at intervals to give a mark for attainment. From the inception of the teacher-training system in the 1840s students had been tested in prose reading by Her Majesty's Inspectors on their annual visitations and marks had been awarded. Reading aloud is still a popular classroom activity.

It is thus easy to understand why the Examining Boards included a test of prose reading in their new spoken English examinations. Reading aloud is a good test of a candidate's ability to interpret and present the ideas and words of others. What was also needed was a test of his ability to present his own ideas and words. Thus Conversation came to be chosen as the second item of a ten-minute test. The J.M.B. developed these forms of tests at 'A' level in the early 1950s, the University of Durham Examinations Board used them for several years until its demise in 1964, London has used them for its 'O' level examinations since their inception in 1964, and they are by far the most popular forms of spoken English test with the new C.S.E. authorities, which instituted examinations in 1965 and 1966.

A test composed of these two items has certain advantages. They are nicely balanced with a contrast of oral interpretation and oral composition. They make for a pleasant variety, and together they test a candidate's ability to communicate ideas and feelings to others. They are easy to administer and need not take more than ten to twelve minutes per candidate. Research has shown that in the hands of competent examiners they have a reasonable statistical validity and reliability.

These tests in Reading and Conversation are, in general, private affairs between the single candidate and the examiner, with no other people present. The Boards have been experimenting for some time with examinations in a group situation and, at the same time, with items other than Reading and Conversation. Conversation is, in its nature, talk between two persons. If three or more take part its form and nature are subtly changed; it becomes Discussion. For the last two years Conversation in the J.M.B. tests has been a three-handed affair - it has become a discussion (on any subject chosen by the candidates) involving two candidates and the examiner. The London Board has also been experimenting and proposes to introduce in 1971 (in addition to its single-candidate Conversation) a Discussion involving three candidates and an examiner. In these two examinations all candidates in any one discussion are being tested. In the 13 regional C.S.E. areas Group Discussion is a compulsory part of the oral English test in two areas and optional in five (which means that candidates can take some other option if they wish). Of the seven syllabuses involved two have group discussion in which all candidates are being tested; in the others each candidate is tested separately, talking with the group.

The London Board is retaining Reading in its new 1971 examination, but as a group activity, each candidate being required to read aloud to a group comprised of the examiner and other candidates. In the J.M.B. tests Reading is now also a group activity, but optional to the giving of a Talk. A new feature is that the candidate sits after his reading and answers questions from members of the group on any matter arising. Reading aloud

is still a very important test item in the C.S.E. examinations, appearing in 12 of the 13 syllabuses, sometimes compulsory, sometimes optional; sometimes a private affair between candidate and examiner, sometimes in the presence of a group of other candidates. In three regional areas the reading tests are conducted with the candidate sitting at a table close to the examiner, in others he stands and speaks at a distance; in one or two areas sitting or standing is at the choice of the candidate. (The South East region also tape-records its candidates for purposes of moderation.)

The size of group for the 'A' level J.M.B. tests is six or seven, for the 'O' level London three. This means that each member of a group spends much more time in his examination than if he were tested solo - in the former test one and a half hours (instead of 12-15 minutes), in the latter half an hour (instead of ten minutes). This is excellent for the generation of a group rapport. The size of the C.S.E. groups varies. In the Metropolitan region Group Test the size is twelve. (This test is unique. The candidate introduces the passage to the other eleven in the group, answers questions on it, and then reads it to the group a second time, two marks being given, one for either reading, and one for the quality of his answers.)

The last important development is the institution of the Talk as a test-item. In the J.M.B. tests it is optional to Reading and is delivered to an audience of six or seven (the rest of the group and the examiner). Questions and answers from the candidate follow. The candidate is handed a printed card containing three topics, from which he chooses one. He is allowed a few minutes to prepare his talk and five minutes to deliver it. He can speak from notes. His audience then questions him on matters arising. The Talk is a compulsory element in the C.S.E. test in four regions and optional in four. Where it is compulsory the syllabus states that the candidate will talk on a topic of his own choice. In all areas but one the talk is delivered to an audience. In five of the eight areas, Question and Answer form part of the test.

Thus the tests now in existence comprise Reading, Conversation, Group Discussion, the Talk to an audience. We have moved from the private to the group situation. Tests are coming to be more realistically based; that is to say, they are concerned more with what people actually do in real life situations - they talk to each other singly or in groups, they lead a discussion (or are led), and they stand up and address an audience.

The tests in spoken English most usually taken by candidates for whom English is not the mother-tongue are those for the Cambridge Proficiency Certificates - comprising Prose Reading and Conversation with the examiner. Among the speech elements the examiner is concerned with in those tests are those that help him to answer the very important question - how English does this candidate sound? They are his pronunciation (his use of vowels and consonants), his intonation (or tune-patterns), his articulation, and, perhaps most important of all, his pattern of stressing, a complicated alternation of stressed and unstressed syllables. These are the major elements in the English speech rhythm. In Conversation the examiner is also looking for the use of an acceptable vocabulary, acceptable grammar, word-order and idiom.

Now the English candidate, however poor his speaking in other respects, at least sounds English - he can't help it. So the examiner in the tests in England under discussion is not primarily listening for the way

these speech elements are used. In Reading the examiner is looking for the candidate's ability to interpret the page before him, to communicate to the audience the author's meaning and mood. He is asking the candidate to exercise his imagination as well as his communication skills of voice and speech. In the other test elements - conversation, discussion, the talk - the examiner is considering the candidate's ability to use the English Language efficiently in face-to-face communication - that is, to make a statement clearly, to develop a theme, to rebut an argument, to inform, to persuade. He is, of course, also considering the various aspects of delivery (the use of the voice, dictions, bodily stance and gesture).

It will be seen that judgments made by an assessor about a candidate's speaking are necessarily highly subjective - he has to make the decision as to whether a speaking performance is a good one or a poor one and thus whether to award a high or a low mark. This means that examiners may disagree sharply about the performance of particular candidates. In fact the quality of the examiner is of crucial importance. As much as possible is done by test-designers, by the examining body and by chief examiners to minimise the possibility of disagreement. This is done by "standardising". A rating scale is prepared which shows what qualities are being looked for in a candidate's speech and what mark is to be awarded for the strongly positive presence of a quality and what for the almost total absence of this quality (e.g. CONTENT and ORGANISATION OF TALK: Main points made clearly in a logically developing argument. Content clearly organised to show an introduction, a middle and a conclusion. Material interesting, relevant, sufficient, of good quality - AWARD 7 to 10 marks. Talk badly arranged. No logical development or argument: main points do not stand out clearly. Material of poor quality, uninteresting, irrelevant, insufficient - AWARD 0 to 3 marks). Then there will be a briefing meeting at which the Chief Examiner will take the assistant examiners carefully through the rating scale so as to establish an identity of understanding as to what is intended. It is probable that a few "guinea-pig" candidates will be examined by the Chief and the other examiners so that the latter can have a preliminary experience of both examining and marking, and so that standards shall be set and absorbed. When the examiners are in the field examining the Chief will pay each a visit for a day or half a day and make his own assessments of the candidates. Later these will all be scrutinised and the assistants' assessments raised or lowered or left as they are. It is a chancy business. But research has suggested that the judgments of experienced examiners in spoken English (at least in Reading, Conversation and the Talk) are at least as valid and reliable as those of written essay-type examinations. (There is considerable doubt about the reliability of group discussion assessments).

If examinations (over the whole range of educational activities in school) are to remain as a vital element in assessment it seems certain that the assessment of spoken English will become a "growth" industry. Perhaps the most important problem will concern the calibre and training of potential assessors. (In the C.S.E. testing of spoken English almost all teachers of English are likely to be drawn into assessment.) There will also be the search for new and appropriate test-elements that test achievement in school courses in spoken English, which must themselves be geared to the needs of human beings in adult society. Finally, there will be continuing research into methods of assessment and their attendant problems of validity and reliability.

IMPROVING PRACTICAL EXAMINATIONS IN SCIENCE SUBJECTS*

R.H. Dave and Y.B. Patwardhan

National Institute of Education,
National Centre for Educational Research and Training, New Delhi, India.

Examination reform has now been accepted in our country as a very powerful instrument to improve quality in education. During the implementation of the reform programme particularly in science subjects it was felt that the work in these subjects would be incomplete unless the practical examinations are also reformed. In fact, the sixth conference of chairmen and secretaries of the boards of secondary education in the country held in November 1964 discussed at length the necessity of improving practical examinations and passed a resolution that the boards may take up this work in collaboration with the Central Examination Unit of the NCERT.

It was a source of great satisfaction that soon after this conference, the Board of Secondary Education, Rajasthan, came forward to take up the new venture in the field of examination reform. A series of experimental tryouts were carried out to evolve an improved system of practical examination and then the board implemented the new plan in its higher secondary examination of 1968 in the subjects of physics, chemistry and biology after making adequate preparation in collaboration with the NCERT. We give a brief résumé of the experimental studies conducted for the development of an improved pattern of practical examinations in science subjects and its large scale implementation.

SHORTCOMINGS OF THE PRESENT PATTERN

A qualitative study based on verbal reports by a number of experienced examiners in Rajasthan revealed that the then existing pattern of examination suffered from the following major shortcomings:

1. Poor sampling: Each experiment given being very complex, comprehensive, and time consuming, only a few experiments (e.g., two in physics) could be set in the limited time available. As such it could only measure a small fragment of the content and a few of the many aspects of skill that practical work is expected to develop. This reduced both the validity and reliability of the practical examination. This was very discouraging to the pupil as well as to the teacher, especially on account of a high degree of chance factor operating in such a system.

2. Absence of reliable criteria of assessment: The criteria of assessment were very general and examiners were inclined to assess the performance of students according to varied standards leading to loss of consistency and uniformity.

---

* This article is reprinted, with permission, from the NIE Journal, September 1969.
(Publication Unit, NIE Campus, Sri Aurobindo Marg, New Delhi 16, India).

3. Non-comparability of exercises: The few exercises (e.g., two in physics) each pupil would get, varied very much in complexity and nature of skill involved. It was not justifiable to compare the performance of different candidates as obtained on these differing instruments.

BASES FOR A NEW PATTERN

It was felt that practical tests are more costly and time-consuming, and so they should be used only when other more convenient techniques such as written tests cannot be used. It is in the realm of practical skills in these subjects that written tests are not usable and hence practical tests should essentially be used to measure practical skill although other objectives such as knowledge, understanding or application need not be entirely eliminated. For this purpose practical skill was defined under the heads:

(a) process of performance, and

(b) product of performance.

They were further clarified in each subject to delimit their scope. In physics, for example, they were delimited as follows:

Process of Performance

The pupil

1. selects appropriate apparatus, tools, etc.
2. checks apparatus, tools, regarding their working.
3. detects errors and limitations in the fitting up of apparatus.
4. rectifies errors, if possible, under laboratory situation.
5. cleans apparatus, tools, etc.
6. sets up apparatus, tools, etc.
7. sketches arrangement of apparatus (if necessary, at the outset).
8. prepares and follows a systematic and sequential plan for taking observations.
9. states the principle, formula (explaining the symbols, etc., useful in the experiment).
10. manipulates apparatus, tools, etc., while performing the experiment.
11. measures quantities and reads instruments, apparatus, etc., accurately.
12. takes precautions in handling instruments, substances, etc.
13. makes accurate observations of parts, specimens, processes, etc.
14. records observations and makes calculations where necessary.
15. verifies observations.
16. performs experiments with reasonable speed.
17. performs experiments with reasonable accuracy.
18. performs experiments with neatness.
19. adapts himself with somewhat new and different apparatus in setting novel experiments.
20. explains orally the procedures, principles, etc., involved in the experiments.

Product of Performance

The pupil

1. summarizes observations.
2. calculates and finalizes the results.
3. interprets data and draws conclusions.
4. records experimental procedure and conclusions.
5. dismantles and cleans the apparatus, where necessary.
6. arranges the apparatus, substances, etc., at their appropriate places at the end of the work.

Sessional Practical Work

The practical exercises performed by pupils in the higher secondary classes are recorded in specially developed record books. The skills and traits attained while performing these may also be evaluated in board examinations. It may not be possible and also not desirable to evaluate all the traits developed, but a few like completeness, neatness and regularity may be evaluated with the assistance of the subject teacher.

Some of the skills that may be appraised from this aspect of practical work are:

1. Drawing diagrams and sketches from observed facts.
2. Collecting specimens like that of ores, minerals, crystals, etc.
3. Displaying material collected.
4. Improvising simple apparatus.
5. Constructing models.

DEVELOPMENT OF A NEW PATTERN

For the purpose of improving validity and reliability the practical examinations were modified in the following respects:

1. Increasing the number of exercises: Instead of giving few long exercises, many short exercises are introduced, e.g., in physics one major comprehensive experiment is retained and the other is replaced by four or five short exercises. The maximum marks and the time, however, are kept the same.
2. Making the exercises objective-based: Exercises are to be set to test predetermined specific aspects of skill (or understanding) as laid down in the objectives. As mentioned earlier they include within the process and the product of performance. This tends to improve the validity of the practical examination.
3. Improving the sampling of abilities and content: Increase in the number of exercises enables the test to cover many different abilities as specified under the specifications of the skill objective and also to cover a variety of content areas. This helps in improving the coverage and consequently tends to improve reliability and validity.
4. Improving scoring procedures: Very detailed marking schemes giving minute analytical details of assessment of pupil performance are developed not only for major and short

exercises but also for sessional work and viva. Detailed instructions are developed for the use of examiners and candidates for this purpose. This helps in improving objectivity of scoring and controlling inter-rater reliability by minimizing the variability in scoring by examiners emerging from extraneous factors like personal likes and dislikes.

5. Improving reporting and interpreting procedures: Detailed proformas and instructions for their use are developed for the use of examiners. When these reports would be properly used by schools, they will be able to improve science teaching in many respects.

TRY-OUTS

Four examiners in each subjects of physics, chemistry and biology who were involved in the development of the new pattern tried out these procedures three times in actual situation specially arranged for this purpose. In the first try-out 10 candidates and in the second and third try-outs 15 to 20 candidates were involved in all the three subjects. In all the try-outs the four examiners observed simultaneously and marked independently. The experience of earlier try-out was always invariably used to improve the exercises and refine the scoring schemes of the subsequent try-outs. The results of assessment were then studied and correlations found. The findings in the try-outs of biology are given here. In other subjects similar findings are available.

Comparative Study of Three Sets of Examiner Inter-correlations in Biology Practical Examination in Three Try-outs

A. Averaged across questions

Examiner/Try-out	AB	AC	AD	BC	BD	CD	N
I.	.78	.65	.67	.83	.74	.68	12
II.	.94	.95	.95	.92	.94	.93	15
III.	1.00	1.00	1.00	.99	.99	.99	15

B. Based on total test scores

Examiner/Try-out	AB	AC	AD	BC	BD	CD	N
I.	.89	.88	.59	.78	.82	.63	12
II.	.87	.45	.84	.55	.90	.45	15
III.	.97	.99	.98	.98	.98	.97	15

C. Based on ranking of difficulty indices of questions

Examiner/Try-out	AB	AC	AD	BC	BD	CD	N
I.	.88	.90	.90	.98	.78	.75	9
II.	.93	1.00	.98	.93	.97	.98	9
III.	.98	1.00	1.00	.98	.98	1.00	9

Significance levels of Rhos:

When N = 9: .05 = .600; .01 = .783
When N = 12: .05 = .506; .01 = .712
and When N = 15: .05 = .439; .01 = .623

In this subject, during the first try-out three out of six examiner inter-correlations are not significant at the .01 level but are significant at the .05 level. The agreement among the examiners in Try-out II is at or above .92. All the inter-correlations are substantially increased in Try-out II. Again, in Try-out III, the examiner inter-correlations averaged across the questions have reached unity in 3 out of 6 cases, and in the rest they are at .99. Thus the inter-examiner agreement in biology practical tests reached almost to the optimum as a result of intensive training, practice and development of well-designed scoring procedures.

IMPLEMENTATION

1. Preparation: Encouraged by the above findings, the Rajasthan Board decided to launch the reform programme on a large scale throughout the state. It, therefore, developed brochures in each subject entitled "Improved Pattern of Practical Examination" with the help of the NCERT and the examiners who participated in the try-outs, and circulated them to schools. The board also trained all examiners in new pattern of practical examination in the three science subjects. They were given training in the theory of conducting the examinations through four-day workshops organized for the purpose. They acquired practical experience in conducting the new type practical tests in actual examination, which were specially arranged at various places as a part of the training programme.
2. Implementation: With these and other preparatory steps carefully executed, the board introduced the new pattern in the higher secondary examination of 1968. Its impact on school practices is being closely watched. The preliminary review of the impact has been found to be quite encouraging.

SOME PROBLEMS

During implementation some problems were faced which were already envisaged.

1. Lack of equipped laboratories: Many schools do not have good laboratories. For want of such laboratories it becomes difficult to conduct the examination effectively. This applies to the old pattern of examination also.

   Some laboratories do not have trained assistants. Services of trained assistants are essential.
2. Number of candidates per examiner: This pattern envisages close observation of pupil performance during the period of examination. One examiner cannot obviously cope up with 20 candidates at a time as is the practice in vogue. Perhaps, 10 may be a manageable number.
3. Trained examiners: For some years, till the examiners are acquainted with the new pattern, it will have to be seen that every examiner is fully acquainted with the spirit and technique of the new pattern of examination before he is entrusted with the job.

IMPLICATION

For the efficacy of this new examination co-operation from different agencies will be needed. Some implications to such agencies are indicated below:

Departments of Education and Boards of Secondary Education

1. School laboratories will have to be better equipped.
2. Practical syllabus may be reviewed.
3. Flexible time-tables will have to be permitted.
4. Better inspection and guidance programme will have to be developed.
5. Only qualified and trained examiners will have to be selected.
6. Examiners' reports will have to be scrutinized and the findings reported to schools for action.

Schools

1. More initiative on the part of individual teachers and pupils will be needed.
2. Rigidity of time-tables will have to be reduced.
3. Laboratories should be better equipped.
4. The evaluation data should be used for remediation and improvement.

Teachers

1. Variety of practical activities will have to be designed and organized to develop specific skills among pupils and to discourage the tendency of mechanical repetition of standard experiments.
2. Initiative on the part of pupils should be encouraged.

CONCLUSION

Practical work in science subjects is aimed at the realization of some specific purposes which cannot be otherwise achieved. Practical examinations, therefore, have to be so planned that they measure the degree of success achieved by practical work as a contribution towards the multi-faced development of pupils' innate powers and subsequent achievements. The new pattern suggested here aims at this. It defines the outcomes of the process and the product of performance, stresses the need for developing valid and reliable tools of measurement and builds in ways to evaluate the results of measurement. It also envisages sound feedback procedures to utilize the results of evaluation in improving school practices. It is hoped that given a fair trial this pattern will work as a catalytic agent in making science education a dynamic process and a creative activity in our schools.

THE SOUTH PACIFIC COMMISSION EDUCATIONAL TESTING PROGRAMME

J.W. Taylor

Education Specialist, South Pacific Commission

In common with all developing areas, most of the island territories of the Pacific are faced with financial problems that make universal education a difficult if not impossible goal to achieve at present. In many territories primary education is not compulsory, and in the majority only a limited number can be admitted to secondary schools. For obvious reasons, when only limited numbers are admitted at secondary level, the available resources must be concentrated on those who will contribute most in later years to the development of the territory. Generally speaking, this means those who are intelligent, conscientious, stable pupils for whom there is a good educational and vocational prognosis.

Pupils in academic classes in Pacific secondary schools normally sit metropolitan public examinations (e.g. the Overseas Cambridge Certificate or the New Zealand School Certificate) after three to five years of preparation.

Selection procedures for those proceeding to secondary level in most territories tend to concentrate on the results of academic attainment at upper primary level. Often these tests have not been well constructed, are not representative of the primary school syllabus, are weighted in favour of less important school subjects, and, most important of all, have not been successful in identifying those pupils best suited for further education. School records indicate a heavy drop-out and failure rate; pass rates at metropolitan examination level have been disappointing; and there have not been enough graduates at secondary, tertiary, professional and skilled vocational level to hasten the replacement of expatriates by local appointees at a rate which the optimum territorial development requires.

The crux of the problem, then, appears to be in the selection procedures. Expressions of dissatisfaction and concern resulted in the South Pacific Commission organising a Technical Meeting on Selection and Guidance at Goroka, New Guinea, in 1967. This meeting was attended by senior education officers responsible for selection procedures in Pacific territories. The whole problem was discussed in detail, under the guidance of two consultants skilled in testing and selection principles and practices. As a result of the Meeting, the South Pacific Commission appointed an education specialist who was given the responsibility to assist Pacific territories in the improvement of selection procedures.

In this connection, several basic problems must be faced. In the first place, there is the need to compensate for the unevenness of educational opportunity within each territory - for example the problem of the intelligent child, on some small island, who is culturally deprived and has probably had poor teachers, compared with the child form the administrative centre who may be less intelligent but who has had the advantage of a better cultural environment and better teachers. Secondly, there is the language problem, where the quality of teaching is again uneven, and when many children hear the metropolitan language (which is the language of secondary education)

spoken only in the school - and then it is often of poor quality and limited vocabulary. Thirdly, the length of secondary education must be taken into account. In some schools there is time for remedial work for disadvantaged children before they must face the metropolitan examination - for others the years at secondary school are too few to allow this.

After a careful review of the situation, the decision was made that selection should be based on two criteria - academic attainment in the basic subjects, and academic potential. For preference, both criteria should be in the form of standardised tests, and academic attainment should be based on understanding rather than on mechanical accuracy.

The administration of the selection tests also had to be considered. The Pacific is dotted with small inhabited islands, most of which have at least one school, and many of which have very poor communications. Social pressures weigh heavily on teachers to ensure that children pass examinations. In the important selection tests, therefore, it is necessary to have the tests administered by as few people as possible. Some of those concerned in testing are not trained personnel, and therefore the tests must be easily administered in standardised procedures that are not difficult to follow; and because these testers may be ashore on islands for only a few hours, they must be able to test comparatively large numbers at once. Because of the language problem, the tests of potential need to be non-verbal and as culture fair as possible. Simply administered standardised group tests were therefore indicated.

With these considerations in mind, a large number of existing tests of various kinds were surveyed. Many were found wanting for Pacific conditions. Others seemed more appropriate, were tried, and for various reasons - mainly lack of validity - were rejected. Finally, from experience gained, and from research, a battery of tests was constructed which seemed to cope with the requirements of the situation. The battery comprises the following:

(1) Tests of academic potential:

(a) A test of speed and accuracy

This test is the first in the battery. Results are not significant, but the item is retained as a "settling down" item for subjects, to enable them to overcome nervousness, to direct their attention to detail on the printed page, and to make them aware that time is an element in the test. (In later tests, concentration is on power, the time limit being adequate for most subjects to complete those items within their ability.) Test validity in pilot surveys was increased when this item was left in.

(b) Two reasoning tests based on symbolic material

These comprise series of numbers or letter in logical sequence, the subject being required to give the missing item in the series.

(c) A figure-grouping test

This is entirely non-verbal, and of this type of test, correlates best with academic success in the pilot surveys.

(d) A test of general ability

An adaptation of the Papua-New Guinea Reasoning Series.

(2) Tests of academic attainment

(e) A test of mathematical concepts

This test is based on understanding of the new approach to mathematics. It can be combined, where desired, with a mechanical accuracy test.

(f) A test of English

Still under revision, and based on the vocabulary and structures taught in the widely used Oral English syllabus produced by the SPC Language Specialist, Miss G. Tate.

(g) A test of reading and comprehension

Under construction. Will include word knowledge, speed of reading, general comprehension.

In the selection procedure the results of the tests, reduced to standard scores, may then be distributed on a scattergram, with the mark for academic performance on one axis, and the mark for academic potential on the other. The small number tested make this a feasible method in most territories. Selection procedures are then facilitated, and can be adjusted according to the educational system, and the number of places available in the secondary system. For example, territories with a short span of years before the metropolitan examination would select only those who score well on both criteria; while those with a longer span may include those with high potential and low performance on the basis that there is time for remedial work in the initial years at secondary school.

With these tests, it is recommended that teacher opinion be taken into account (especially in marginal cases) when making final selections. This should be as objective as possible, through the use of a checklist questionnaire which takes into account personality and social factors, attitudes towards study, conscientiousness etc..

The above tests have been used in pilot surveys in 1969. They are now being subjected to statistical checks, to be produced in final form for standardisation and validation in the immediate future. In addition to these tests, after experience with New Zealand Maori pupils and one group of Pacific Islanders, the Raven's Matrices Test is being investigated for use in the Pacific. Particular attention, on the basis of controlled research, is being given to the possibility of improved validity when the Coloured Matrices Test is administered as a practice item some days before the full Matrices

test is used as the test item. Initial results seem to indicate a higher validity with this procedure than with the administration of the full test alone.

The larger territories of the Pacific have developed their own testing services. The South Pacific Commission, for example, was able to draw on the experience of the Psychological Service of Papua-New Guinea; while in Fiji the new University of the South Pacific is directly involved in selection procedures at several levels of education. The South Pacific Commission service is directed mainly at assisting the smaller territories which cannot afford these specialist services, and have not adequate resources to cope with the problem internally. It is hoped to publish a full account of these investigations when final standardisations and validity have been established.

THE EAST AFRICAN EXAMINATIONS COUNCIL

B.P. Kiwanuka

Registrar, East African Examinations Council, Kampala, Uganda.

Historical Background

The East African Examinations Council was established towards the end of 1967 by an Act of the East African Legislative Assembly. The Act is generally referred to as the East African Examinations Council Act, 1967. It was enacted on behalf of the East African Common Services Organisation which later became the East African Community. The Community includes the three States of East Africa, namely Kenya, Uganda and Tanzania. It is responsible for running services that are common to the three States, e.g. Railways, Airways, the Posts and Telecommunications, School Examinations, etc.

Up to 1967 all the major school leaving examinations were conducted by examining bodies based in Britain. These included the University of Cambridge Local Examinations Syndicate, the University of London School Examinations Council, the Associated Examinations Board, the City and Guilds of London Institute, the Royal Society of Arts and a few others which concentrated on specialised fields. The Council Act of 1967 was a climax of a move which started in the early sixties. In 1964 the Creaser Committee, which was appointed to look into University Entry Requirements, reported, among other things, that

> "Over the last two years the Academic Committee of the Provisional Council of the University of East Africa and its successor, the Senate, has been concerned that entrance to degree courses within the University should relate to national needs ..... There is strong pressure for the early establishment of an East African Examinations Council to take over from the Cambridge Syndicate the external school examining at the form 4 and the form 6 levels. Such a step can be justified on both educational and political grounds."

In 1965 the Vice-Chancellor of the University of East Africa invited the Cambridge Syndicate to send an adviser to East Africa to consult with the Governments and the University on the possibility of setting up an examinations Council. As a result of this invitation Mr. A.V. Hardy, Deputy Secretary of the Cambridge Syndicate, came to East Africa and held discussions with representatives of the Governments and the University. He then prepared a report outlining ways in which an Examinations Council could be formed. It was on the basis of this report that the Council was established by the Act referred to above.

The Act provides for the representation of the main bodies concerned with education in East Africa on the Council, namely the Governments of the partner States (Kenya, Uganda and Tanzania), the East African Community, the University of East Africa and its constituent Colleges, the Heads of Schools and the Teachers. It specifies that "The objects of the Council shall be to conduct within East Africa such academic, technical and other

examinations as the Council may consider necessary or desirable in the public interest."

The Committee Structure of the Council:

The steering body of the Council is called the Finance and General Purposes Committee. This is, in effect, the Executive Body which supervises the implementation of the Council decisions. As in the case of the Council, the Finance and General Purposes Committee membership reflects the interests of the Partner States, the East African Community and the University.

In addition to this Committee, the Council is empowered to appoint, and has actually appointed, other Committees for specific purposes. There are, for instance, National Sub-Committees (one in each Partner State) whose function is to consider the work of the Council in relation to the specific requirements of their respective countries. They advise their respective Governments on the subjects that should be examined by the Council and they comment on the examinations and examination papers taken each year by candidates in their respective countries. There is a School Examinations Committee whose main function is to advise the Council on the suitability of new syllabuses, the suitability of examiners to be appointed and the regulations to be used. There are also various International and National Subject Panels covering all subjects examined. The purpose of these panels is to study existing syllabuses and improve on them as necessary, develop new syllabuses, delete unnecessary ones and then present all these to the School Examinations Committee for approval on behalf of the Council. They also make recommendations on suitable examination setters.

The Committee Structure of the Council emphasizes the need for considerable consultation and involvement by all concerned in this venture. Action can, for instance, be initiated by an individual teacher and is passed upwards through the National Subject Panels, the National Committee or the International Subject Panel to the School Examinations Committee or the Council itself. Proposals initiated at the Council level or the School Examinations Committee level have a way of reaching the teacher in the school in each of the participating countries. In this way it is hoped that the East African Examinations Council will be a people's Council and will not be looked at as a mysterious body that imposes syllabuses or examinations on countries and schools without their active participation.

Council Activities:

The scope of Council's activities, as set out in the Council Act, is fairly wide. The Council has therefore decided to be a little cautious in its approach to its task in order to ensure that the foundation is well done. Initially it was decided to concentrate on the secondary leaving examinations. As reported above there already existed a demand for participation in these examinations.

In 1968 the Council and the Cambridge Syndicate agreed on a programme whereby the Council would gradually take over the examining activities for both the O-level and the A-level secondary school leaving examinations. The Syndicate would gradually phase out. It was decided that a first step in this exercise would be the joint awarding of certificates by the two bodies. Hence the former Cambridge School Certificate/G.C.E. and the Higher School Certificate have been replaced by the East African Certificate of Education/East African School Certificate and the East African Advanced

Certificate of Education awarded jointly by the Council and Syndicate.

This joint operation, in the transitional period, is also reflected in the actual setting and marking of examination papers. For the next five years or so papers for the O-level and A-level examinations will be set and marked partly in East Africa and partly in Cambridge. The take over programme includes the training of East African examiners by Cambridge Examiners. The East African Examiners are recruited by the Council through the Ministries of Education of the three countries. They are trained in East Africa by Instructors recruited by Cambridge and those who are considered suitable are appointed as Examiners in their respective subjects.

The recruitment of suitable examiners is one of the major challenges to the Council. Obviously large numbers of examiners are required if the Council is to take over responsibility for all subjects. There are now about 35,000 candidates at the O-level and 4,000 to 5,000 candidates at the A-level. There is always the big problem of the wastage of trained examiners and this is aggravated in East Africa by the fact that a large number of teachers are still expatriates. The majority of these come to work in East Africa for short contract periods. Their assistance in this exercise can only be of a temporary nature. However, on the lighter side of the problem, there is the determination of each of the participating countries to train its own local teachers and reduce the reliance on teachers from abroad. It is thus hoped that this will not only provide the Council with the examiners required, but will also provide continuity which is essential for an examination system.

Reference has been made above to the establishment of subject panels both at the national level and the international level. The Council considers it its obligation to assist the partner States in their effort to localise syllabuses. Already some of the subject syllabuses have been tailored to suit the requirements of the member countries. Examples of these include syllabuses for History, Geography, Physics Chemistry, Biology and Literature in English. The moulding of syllabuses is initially done by the National Panels for the particular subject and the final versions are agreed upon at meetings of the International subject panels.

The revision or development of new syllabuses does, of course, bring its own new problems. First, one has to consider not only the relevancy of the matter included in the syllabus, but also its standard in comparison to existing syllabuses. Subject panels must therefore ensure that the revision or replacement of syllabuses does not result in the lowering of standards. Another problem related to that matter is the designing of new examination papers based on the new syllabuses. Problems of this nature have been experienced in connection with the development of School Science Project syllabuses. It was discovered that because of the different rates of development between Chemistry and Physics, it was impossible to set papers of the same standard for all candidates in the three countries. The solution to the problem was to set papers including alternatives of the new and old syllabuses - the so-called hybrids. This is, of course, an interim measure which will disappear when all syllabuses have been fully developed.

This paper is being written shortly after the first cycle of the examining exercise has been completed. This cycle has been an eye opener to the Council. Reference has been made to the processes of recruiting and training examiners. The exercise has revealed problems of security which, for obvious reasons, cannot be discussed in a paper of this nature. The problems of long distances in a broad region like East Africa have also come

to light. Whereas in Britain, and possibly in other countries, marking can be done in the examiners' homes, in East Africa all the marking must be residential. This certainly adds to the cost of the exercise. But it reduces the chances of losing candidates' scripts on one hand, and on the other hand it makes coordination and standardization easy. Despite all the problems that may crop up in such a venture the staff of the Council and the Council itself have been greatly encouraged by the co-operation of people within East Africa and friends outside East Africa.

GUIDANCE AND COUNSELLING IN MALAYSIAN SECONDARY SCHOOLS

G. Krishna Iyer

Chief Guidance Officer, Ministry of Education, Malaysia.

The educational system operative in West Malaysia provides for six years of free primary education available in the four media of Malay, English, Chinese, and Tamil. A further three years of education is available in Malay and English. These three years of education comprise the comprehensive stage of education; subsequently the pupils are streamed for general, vocational and technical education. Pupils who proceed for post-secondary education are selected on the basis of their academic performance. There is a great demand for education and a consequent need for guidance to help pupils make intelligent choices and adjustments so that in the long term they will benefit from the education provided, choose their own way of life and attain vocational and life goals satisfactory to the individual and of relevance to the country in its present state of development.

Before discussing the programming of the guidance system in secondary schools it is pertinent to view those areas in the life of the Malaysian adolescent which tend to pose problems with which they need help. Some of the problems posed are those common to adolescents the world over, others are peculiar to the youth of the Malaysian cultural milieu.

One major problem area may be termed adjustmental. The adolescent pupil has to adjust himself to being taught in a medium different from that he was used to during his primary stage of education. He has to adjust and integrate himself to the less personal secondary school. He becomes increasingly aware of cultural and social differences amongst his peers. He is posed with the problem of choosing his electives soon after his entrance into the secondary school. All these problems assail him, when he is on the brink of adolescence and when the individual is in a new environment, and in need of help to orient himself to the new roles expected of him.

Another major problem area may be described as developmental. The secondary school adolescent finds that he has to develop his traits in certain directions if he is to be regarded as developing positively in the eyes of critical adults. In West Malaysia, as elsewhere, a horde of influences impinge on the impressionable adolescent. He often perceives a sharp dichotomy between expression and implementation. There is also the generation gap in attitudes and values between the adolescent and his parents, and he finds difficulty in reconciling the preachings of adults to the situation in which he finds himself. Rapid changes in the social, economic and political conditions in the country aggravate this problem area.

As he approaches the end of his formal secondary school-life the problem of choice of career looms large. In an essentially agricultural country with a large youth population, where industrialization is relatively at its infancy, the world of work is highly competitive and selective. In a situation where potential labour is in excess of jobs available, the average adolescent pupil is in dire need of guidance and counselling. The realities of the harsh world of adults are realities which the adolescent has to under-

stand and adjust to in a relatively short time if he is to eke out a satisfactory livelihood.

In summary it might be said that the overall problem of the secondary pupil is this: that of developing from adolescence to adulthood and at the same time overcoming handicaps or adverse conditions simultaneously in a relatively short time.

It will be seen that there is a need for a guidance and counselling service in Malaysian secondary schools. This need has been recognised by the Ministry of Education, and through its direction, guidance is becoming an increasingly important facet of secondary education.

Admittedly guidance is a relatively new approach in secondary schools, but a fair amount of activity is being undertaken by schools in order to alleviate the problem areas in the life of adolescents.

The Ministry of Education has to date undertaken the task of planning and organising Guidance in schools. Some of the preliminary work in this connection has included the introduction and maintenance of cumulative record cards, the compilation of bulletins on guidance for the information and implementation of the guidance teacher, follow-up studies of guidance in schools, and the organisation of seminars and in-service courses for guidance teachers. Shortly, six officers are to be appointed as Regional Guidance Officers to ensure that effective guidance is available in schools. These various steps are all geared to provide guidance at schools with the boost it deserves.

In the secondary school itself guidance is provided in a variety of ways. The guidance teacher is responsible for the guidance programme in each school, but the guidance function is carried out by all teachers. In the bigger residential secondary schools the house system has been used not only for games participation but also for the development of the tutorial tradition where a teacher is a mentor and friend to a group of pupils who turn to him for guidance in matters of personal development, social relations, learning difficulties and allied problems of adolescents. In other schools the guidance teacher is especially allotted a timetable comprising in the main the teaching of Civics. Thus the teacher has that much more time, while teaching Civics, to undertake group guidance, discuss values and attitudes, teach occupations and prepare pupils for adult roles. In some other schools the time allotted for the teaching of Home Science is partly devoted to guiding pupils in matters of code of conduct, personal difficulties, and the establishment of satisfactory relationships with the other sex. The secondary school curriculum provides an excellent springboard for an indirect guidance role.

The guidance programme in a secondary school may be described as geared to the purposes of orientation, appraisal of the individual pupil, conveyance of job information, and helping pupils in personal-social relations. Undoubtedly the achievement of these objectives requires the co-operation of not only the teaching staff, but also the expertise and active support of other agencies, both governmental and private bodies.

Throughout the secondary school career of an individual pupil a personality profile of the individual emerges through the maintenance of the cumulative record card, anecdotal records of him, record of his test results, his activities membership, and the record of his interviews with the guidance teacher. On arrival at the first year of secondary schooling, he is

familiarised to his new environment, through an orientation programme. A pupil who is weak in certain subject might be given special attention. All pupils have access to the guidance teacher, and on occasion the guidance teacher might make home visits to gain the co-operation of parents in sorting out problems.

In the second and third years, while the guidance teacher organises the programme to place sufficient stress on the personal and educational development of pupils, the main stress is laid on helping pupils in their career development. For this purpose pupils are helped in their choice of electives, the teacher of Civics or English teaches pupils in areas related to the world of work, such as on how to apply for a job, the nature of job interviews, the particular careers open to pupils with various stages of secondary schooling, and he encourages pupils to think and articulate on themes related to their ambition or their self-concept. The pupils are helped to explore their interests and aptitudes by making available to them experiences which are job-oriented. Thus pupils are encouraged to think about the job implications of their participation in extra-mural activities. Organised visits to places of training and employment, the viewing of films on jobs, the provision of job experiences, and group project work in the detailed study of certain occupations are all primed to the broad vocational development of pupils and to the aim of helping them to specify in their own minds on a rational basis the career of their choice.

In the fourth, fifth, sixth and seventh year the programme outlined above is continued with certain modifications brought about by the fact that the pupils are older, more mature, and the problem areas are those of a personal nature or those related to entry into the world of work or further studies. The persons concerned with guidance in schools have liaison with personnel or persons from the Ministries of Labour, Health and Social Welfare, philanthropic organisations such as Apex, Rotary and the Council of Child Welfare.

Much is being done in the area of guidance, but counselling proper is in the main absent. True the guidance teacher does help pupils through talking to them, interviewing them and suggesting courses of action they might take, but counselling by properly trained personnel exists only in a couple of secondary schools.

It is recognised that it would be desirable to have at least one full-time guidance teacher in each secondary school but at the moment this cannot be implemented in full because there are other priorities to be considered as well. However there is a growing number of advocates of guidance, and the crucial factor in this respect might well be the assessment of the effectiveness of guidance by the principals of the school. In the final analysis their testimony on guidance in the secondary schools might well give a new boost to the guidance course.

DEVELOPMENT OF THE CANADIAN SCHOLASTIC APTITUDE TEST

Ross E. Traub
The Ontario Institute for Studies in Education

and

H.A. Elliott
Service for Admission to College and University

The Canadian Scholastic Aptitude Test (CSAT) for English-speaking students is one of a battery of objectively scorable tests being developed by the Service for Admission to College and University (SACU). CSAT is designed to measure what may be described as the general verbal and mathematical abilities of students in their final year of secondary school. The test is divided into four sections each requiring 30 minutes of administration time. Two sections test verbal ability, one with antonym, verbal analogy and sentence completion items, the other with questions on the content of several short essays dealing with diverse topics. The remaining two sections of the test each contain a heterogeneous collection of mathematical items which assess the examinee's ability to reason numerically, algebraically, or geometrically. In addition to these four sections, CSAT contains a fifth section composed of items being pretested. Performance on this section, which also requires a half-hour of administration time, does not count towards the scores an examinee achieves on the test. The motivation of examinees to perform pretest questions is maintained by concealing their identity in the test.

CSAT was first given in February, 1969 and is presently scheduled for administration once per year. The test is new each year in the sense that it consists of a different set of items drawn from a secure item pool. It should be noted that different forms of CSAT can be made very similar in the sense that each form can be built from items drawn to match similar specifications with respect to type, difficulty and level of discrimination. The specifications are laid down by a test development committee consisting of one member from each Canadian province plus representatives of SACU and of the Ontario Institute for Studies in Education, the institution responsible for item development and test assembly under contract with SACU.

Two "raw" scores are derived for each examinee who takes CSAT, a verbal and a mathematical score. Both scores are obtained from a formula in which a fraction of the number of wrong answers is subtracted from the number of correct answers. The penalty is imposed to discourage random guessing on the test. Examinees are fully informed of the penalty and the rationale underlying its use in a handbook they are given to study several weeks before the administration date. The handbook also contains practice items to make examinees better informed of the nature of CSAT.

For the purpose of reporting on the performance of an examinee both to him and to the universities he designates, raw scores are converted to standard scores. The distribution of standard scores has a mean of 500 and a standard deviation of 100. Thus the effective range of CSAT scores is from 200 to 800.

At this point several questions that are frequently asked with reference to CSAT warrant consideration:

1. Why does Canada need a national university admissions testing program? One answer to this question begins by recognizing two facts of contemporary Canadian education, that it is a matter of provincial responsibility and that control of education is becoming decentralized in important respects. Provincial control of education has resulted in the development of interprovincial differences in secondary school programs. The differences are substantial enough to make difficult any direct and meaningful comparison of records of school achievement from different provinces. Decentralization of control has occurred in some provinces to the extent that the teaching and administrative staff of individual schools have sole responsibility for determining curriculum and evaluating student performance. Consequently, it is often difficult to compare in a meaningful way the school records of applicants from different secondary schools in the same province. National admissions tests, such as CSAT, hold forth the hope of providing universities with a valid basis for comparing applicants from different schools and different provinces.

Another reason for university admissions tests is that they help the universities reach early admission decisions. The policy of admitting some applicants several months before they finish secondary school, subject only to the proviso that they successfully complete secondary school, has been forced on Canadian universities by the practice of provincial governments to finance universities on a formula basis and by the practice of most students to apply to more than one university. Under formula financing, universities typically receive government grants in direct proportion to the number of students they have enrolled. To ensure a full first-year enrolment, thereby ensuring full enrolments in succeeding years and qualifying for maximum government grants, a university will admit students in two phases. Many of the applicants admitted in the first phase will decide ultimately to go elsewhere. When a university knows which applicants are not coming, it is able, in the second phase of admissions, to complete its first-year roster by admitting from the pool of remaining candidates. Inasmuch as the first phase of admissions is made in the absence of a completed secondary school record, universities find it advantageous to have the information provided by valid admissions tests to guide their decision making.

2. Why does CSAT attempt to measure verbal and mathematical ability, nothing more nor less? Our first response to this question is that these abilities seem to represent characteristics of considerable practical and theoretical significance. Over the past 60 years psychologists and educators have found them, either separately or in combination, to correlate moderately well with academic accomplishment of many different types. Moreover, verbal and mathematical ability appear with great consistency in factor-analytic studies of academic achievement. They are central to theories such as Vernon's (1961) on the structure of human abilities. One would expect a good test of verbal and mathematical ability to provide scores with considerable relevance for admissions work.

Another response to the second question focuses on the fact that verbal and mathematical scores are only moderately correlated. This implies that a test like CSAT should provide information about two substantially different aspects of an examinee's capabilities. Consequently, the test should enable university admissions officers to judge applicants in terms of the ability most relevant to their proposed programs. Verbal ability should probably receive more weight in comparing applicants who want to study a language or history. On the other hand, mathematical ability would be expected to receive more weight in comparing applicants for work in mathematics, physics or engineering. More than this, the availability of scores on both abilities should enable admitting institutions to counsel students about the advisability of entering one program of study as opposed to another.

3. Why does Canada require its own test? Why can it not use admissions tests prepared elsewhere? It is true that CSAT is very similar to the Scholastic Aptitude Test (SAT) of the College Entrance Examination Board (CEEB) of the United States. The reason for this is not coincidence. In fact, CEEB has been very generous in its provision of assistance to SACU in initiating CSAT. Moreover, the underlying rationale for the two tests is very similar. These facts notwithstanding, there are points that can be made in support of the development of a Canadian test.

One quite obvious point is that the population of examinees for CSAT differs in some respects from the population for SAT. For example, it appears that in order for the test to be ideally suited to Canadian examinees the mathematical items in CSAT must be somewhat more difficult on the average than the mathematical items in SAT. Also, differences between Canadians and Americans in cultural background and in the use of English means that some questions that would be appropriate for use in one country are inappropriate for use in the other.

Another reason for Canada to build its own university admissions tests is that by so doing it retains control of the specifications for the test. If CSAT should prove to be unsatisfactory in certain respects, given its present specifications, it will be possible to make revisions in an attempt to achieve a better instrument. Such revisions would probably be difficult to have incorporated in a test designed primarily for a United States population.

An additional factor which suggests the need for a Canadian test for English-speaking students is the parallel requirement in Canada for a test for Franco-phone students. Such a test, Test d'aptitude générale aux études post-secondaires (TAGEPS), has been developed by the Institut de recherche pédagogique and was also administered for the first time in 1969. The design of TAGEPS is essentially identical with that of CSAT but the items have been produced and validated independently. In the near future, an attempt will be made to equate scores on CSAT, TAGEPS and SAT.

4. What results have been observed to this point? There has been only one administration of CSAT thus far. The test performed well in the sense that satisfactory estimates of internal consistency reliability were achieved. The coefficients for both the verbal and mathematical scores exceeded .90. Moreover, the distributions of scores were as desired, being unimodal and roughly symmetrical and bell-shaped. Satisfactory discrimination among students across a broad range of ability levels appears to have been achieved in that standard scores extended the full range from 200 to 800. What is relatively unknown at the moment is the predictive validity of CSAT. This cannot be determined satisfactorily until those students who took CSAT in 1969 complete at least their first year of university in 1970. However, some indication of validity is available for a test similar to CSAT which was administered in Ontario in 1967 and 1968. For that test, validity coefficients as high as, or higher than, .60 have been observed for some programs in some Ontario universities. The median validity coefficient across all programs in all Ontario universities was unfortunately considerably smaller, about .30. Thus, it is with an air of wary optimism that we await the initial validity results for CSAT itself.

Reference:

Vernon, P.E. The Structure of Human Abilities (2nd ed.), Methuen, London, 1961.

A STUDY OF OPTIONAL QUESTIONS IN EXAMINATIONS

Lindsay D. Mackay

Faculty of Education, Monash University, Melbourne, Australia

The case for inclusion of optional questions in an examination is frequently argued on the grounds that a particular examination is not designed to measure whether a student possesses knowledge or facts, but whether he has developed particular abilities or skills which can be assessed independently of the particular question or questions answered. If the assumption that these skills and abilities can be assessed independently of the particular question answered is accepted, the inclusion of optional questions in a public examination allows the teacher greater freedom to develop these skills and abilities in any of a number of sections of the broad subject area, and it allows greater freedom for the individual student to pursue his interests through independent study.

From a measurement point of view there are considerable difficulties in the use of optional questions in an examination. Many people would argue that the use of optional questions in an extended answer examination adds one further source of variability to the subjectivity and inaccuracy that already exist in extended answer examinations. If a student is permitted to choose which questions he wishes to answer, the basis for comparability of scores is considerably weakened, because different students will answer samples of questions which are not comparable in content, abilities or objectives. As a result, the content validity of the examination differs considerably for different students. If a choice of questions is available, the questions answered by a particular student are likely to provide a limited, and perhaps distorted, sample of that student's achievement in the course, because he will tend to choose those questions he is best prepared to answer, and because the availability of options gives him greater opportunity to use materials prepared by others to produce an answer which does not reflect his ability. Rather than being fairer to all students, as some advocates of optional questions in an examination would argue, the opportunity to choose among optional questions may help the poorly-prepared student considerably more than it helps the well-prepared student.

This report is a brief summary of some of the results of a study of the effects of optional questions in examinations. As part of this study, the marks awarded to optional questions in a number of examinations, ranging from the sciences to the humanities, have been analysed in an attempt to determine -

(i) the extent to which optional questions differ in difficulty,

(ii) the extent to which optional questions and different marker interact,

(iii) the extent to which students of different ability select optional questions of different difficulty, and

(iv) the extent to which optional questions differ in their marker reliability.

(i) Differences in difficulty of optional questions

One problem in estimating the difficulty of optional questions is that the groups of students attempting each optional question differ in ability. In an examination which consists of a compulsory section in addition to optional questions, it is possible to use a student's score on the compulsory section as an index of his ability in the subject being examined. It is then possible to calculate the average score on each optional question for candidates in each range of scores on the compulsory section (i.e. in each ability range), and to use these average scores to estimate the average score that would have been obtained on each optional question if all students sitting for the examination had attempted each optional question. If the examination does not contain a compulsory section, the average score obtained by the students on all other questions attempted on the examination can be used in a similar way, as an index of the student's ability in the subject of the examination.

An indication of the differences in difficulty of combinations of optional questions available to students can be gauged from the results in Table 1. In this table, the differences between the estimated average scores on the least difficult combination of questions available to students are given for six examinations.

Table 1: Differences in average scores between least difficult and most difficult combination of questions available to students in six examinations containing optional Questions.

Examination	No. of students used in the analysis	Choice available	Difference in estimated average mark between least and most difficult question combinations (expressed as a percentage of possible marks on optional questions)
1. Grade 11 Physics	4360	3 out of 5	9.7%
2. Grade 12 Physics	4450	5 out of 7	4.7%
3. A Grade 12 Humanity	2230	5 out of 14	3.5%
4. A Grade 12 Humanity	2150	5 out of 13	3.3%
5. Grade 12 History	7410	3 out of 13	6.5%
6. Under-graduate Physics	190	5 out of 11	10.2%

The entries in the table can be regarded as estimates of the differences in average marks that would have been obtained had the same candidate attempted two different combinations of questions, or had two candidates of "equal ability" attempted different combinations of questions. It is apparent that considerable differences in the results on an examination containing optional questions can arise from differences in the difficulty of the options offered, and that these can produce a difference in average marks obtained by candidates of equal ability of up to 10% of possible marks.

(ii) Interaction of examiner and optional questions

It has long been recognized that different markers mark essay questions to different standards. In the analyses of examination papers in this study it is clear that markers do not mark all questions to a consistently hard or easy standard, and there is evidence of interaction between marker and optional question which results in a marker marking different questions to different standards. An idea of the magnitude of the effect of this can be gauged from the results in Table 2, in which are given the maximum observed differences in the average marks that would be obtained by candidates of "equal ability" who attempted different question combinations, and whose scripts were marked by different markers. Results in the table are based on analyses of Examinations 3, 4 and 5 in Table 1. Examination 3 was marked by 15 markers, Examination 4 by 14 markers and Examination 5 by 31 markers. The system of allocating scripts to markers was such that each marker marked every possible question.

Table 2: Maximum observed difference between average scores obtained by students of "equal ability" as a result of a number of sources of variation.

Source of variation	Maximum observed difference between average scores obtained by students of "equal ability" as a result of this source of variation.		
	Examination 3	Examination 4	Examination 5
	%	%	%
Different questions (irrespective of marker)	3.5	3.3	6.5
Different markers (irrespective of question)	6.8	5.1	9.5
Same questions, marked by different markers	7.8	10.0	23.1
Same markers, marking different questions	10.8	9.6	18.8
Different questions, different markers	13.0	14.1	29.9

It is apparent that there is considerable interaction of question and marker and that two students of "equal ability" would be expected to obtain average marks which differed by up to 30% of possible marks, depending on the optional questions they selected and the particular marker to whom their scripts were assigned. The entries in Table 2 are the observed maximum differences in average scores between students of equal ability due to various sources of systematic variation in the marks awarded to students. Some students would experience differences considerably greater than these average values, and other would experience differences considerably less than the average values.

If a statistical correction had been applied to the marks awarded to different questions by different markers, so that students of "equal ability" would be expected to obtain the same average mark on each question irrespective of the marker who marked the question, then a considerable percentage of the students in the above examinations would have received a different final grading.

(iii) Ability of students attempting different question combinations

In an examination which consists of a compulsory section and optional questions, it is possible to use the compulsory section mark as an index of the ability of students who attempt different combinations of optional questions. As described earlier, a measure of the difficulty of the question can be obtained by estimating the average mark that would have been obtained had all students attempted th question. There is evidence that the most difficult question combinations were chosen by a group of students of significantly lower average score on the compulsory section than the group of students who chose the least difficult question combinations. Some of the results obtained are summarized in Table 3 for Examinations 1 and 6 in Table 1.

Table 3: Differences between average scores on the compulsory sections of two examinations for students attempting the least difficult and most difficult combinations of optional questions.

Examination	Difference in estimated average score of least and most difficult question combinations (as percentage of total score on all optional questions).	Average compulsory section score of students attempting least difficult questions minus average compulsory section score of students attempting most difficult questions (expressed as a percentage of possible marks).
	%	%
1	9.7	+21.5
6	10.2	+ 8.6

(iv) Differences in marker reliability of optional questions

Each of the scripts completed by the 7410 students who sat for a Grade 12 History examination (Examination 5 in Table 1) were independently marked by two of the 31 markers who marked scripts for this examination. The consistency with which the same answer was independently marked by pairs of markers for each of the 13 optional questions on this examination is expressed in three ways in Table 4. Firstly, for each question, the correlation between the marks awarded by two markers to the same answer is given. The mark/remark correlation values in the Table vary from .53 to .67 for the 13 questions on this examination. Secondly, the percentage of variance in the marks awarded to an answer by one marker which is common to the mark awarded by the second marker is given. This percentage represents the percentage of common variance in the marks awarded to the same answer by two markers. The percentage of the variance of scores which is unreliable variance or error variance can be obtained by subtracting the percentage of common variance from 100%. In the Table, the percentage of variance which is common to the two marks awarded to answers on each question varies from 29% to 46%; that is, the percentage of variance of marks which is error variance varies between 71% and 54%. The third measure of consistency in the Table is the average size of the difference of marks awarded to the same answer by two markers. The values obtained range from an average difference of 9.9% of possible marks to an average difference of 12.5% of possible marks.

Table 4: Correlation between marks awarded independently by 2 markers to answers on each of 13 questions on a Grade 12 History Examination.

Question	Correlation between marks awarded independently by 2 markers.	Percentage of variance common to the marks awarded by 2 markers.	Average difference between marks awarded to the same answer by 2 markers (as a percentage of possible marks.)
		%	%
1	.56	31	11.6
2	.57	33	12.5
3	.53	29	10.5
4	.60	36	10.5
5	.67	45	12.3
6	.55	31	10.0
7	.67	46	10.9
8	.65	43	10.5
9	.65	42	11.4
10	.63	40	9.9
11	.65	43	11.3
12	.65	43	10.4
13	.66	44	12.4

Summary

This report briefly summarises some results of a study of the effects of optional questions in six examinations. The results indicate that

(i) Differences in difficulty of different combinations of optional questions available to students can result in average differences of between 3% and 10% of possible marks between students of equal ability who select different question combinations.

(ii) Different markers mark different optional questions to different standards, and, as a result, students of equal ability who answer different questions and whose answers are marked by different markers could be expected to obtain average marks which differ by up to 30% of possible marks.

(iii) The group of students who select the most difficult combinations of optional questions on an examination are of significantly lower ability than students who answer the least difficult question combinations.

(iv) There are considerable differences in the reliability with which different optional questions are marked.

The report has indicated a number of difficulties associated with the use of optional questions in examinations. Further work is currently in progress to investigate other effects of optional questions in examinations.

THE INFLUENCE OF ANXIETY ON SEVERAL MEASURES OF CLASSROOM PERFORMANCE*

Kenneth E. Sinclair,
University of Sydney,
New South Wales, Australia.

The area of research concerned with the influence of anxiety on human learning and performance has significance for both educational practice and psychological theory. Within an educational context it has particular relevance for procedures used in student evaluation and testing. We live today in a highly test conscious culture. Decisions of major consequence to the individual are increasingly being made on the basis of his performance in tests. It is important, therefore, that the various factors that influence test performance be identified and the nature of their influence determined. There is growing evidence that anxiety is a factor of considerable importance in influencing test performance.

Beyond its relevance for educational measurement, research in this area is also contributing directly to a more precise understanding of human learning and performance. Investigators from quite varied backgrounds have carried out research in this area. Behaviorists (Spence and Spence, 1966), neuro-psychologists (Hebb, 1955; Malmo, 1959) and psychologists adopting a more psychoanalytic position (Sarason et al., 1960) have all developed rival theories designed to explain the influence of anxiety on learning and performance. Within an educational context, Sarason's psychoanalytic position has been found to have greatest relevance.

The influence of anxiety on performance in a variety of laboratory tasks is now quite well documented. Laboratory studies have established that the complexity of the task to be performed and the level of stress (usually defined in terms of level of ego-involvement) inhering in the task are two factors, in particular, which must be considered in explaining the influence of anxiety. Anxiety appears to facilitate performance on simple, straightforward tasks where there is little response competition and to interfere with performance on more complex tasks where response competition is likely (Taylor, 1951; Spence and Taylor, 1951; Taylor and Spence, 1952; Montague, 1953; Standish and Champion, 1960). In conditions where ego-involvement is low, a number of studies have found anxiety to be unrelated to performance (Lucas, 1952; Deese, Lazarus and Keenan, 1953; I.G. Sarason, 1957b; Kalish et al., 1958; Nicholson, 1958; Feshbach and Loeb, 1959), although some studies have found that anxiety facilitated performance (I.G. Sarason, 1956, 1957a; Longnecker, 1962). In conditions of high ego-involvement, anxiety has typically been found to interfere with performance (I.G. Sarason, 1956, 1957a; Nicholson, 1958; Harleston, 1962).

---

* This research was supported by a University of Sydney Research Grant. The cooperation of the N.S.W. Department of Education and the principals, staff and students of the Canterbury, Crows Nest and Drummoyne Boys' High Schools is gratefully acknowledged.

This article is reproduced, with permission, from The Australian Journal of Education, Volume 13, Number 3, October 1969.

While these relationships have frequently been demonstrated in relation to laboratory tasks, rather fewer studies have dealt with the question of the relationship between anxiety, task complexity, level of stress, and performance in more naturally occurring situations such as the classroom. Wrightsman (1962), however, in one study, varied level of stress in relation to aptitude test performance. He found no relationship ($r = -.06$) between anxiety and performance in the low ego-involvement condition and a significant negative relationship ($r = -.37$) in the condition of high ego-involvement. While there had been little change in the performance of low anxious (LA) subjects in the two conditions, the performance of high anxious (HA) subjects was reduced by almost one standard deviation by the stress of the instructions.

In a study with college students as subjects, Paul and Eriksen (1964) carried out a similar analysis using a classroom achievement test. A regular psychology class examination was administered on the morning of the experiment (the high stress condition) and a parallel form of the test was administered to the same individuals at night under conditions designed to minimise anxiety (the low stress condition). When their data were analysed using only subjects from the middle range of intelligence, a significant interaction was found between level of stress and level of anxiety. In the high stress condition, LA subjects were superior to HA subjects, while, in the relaxed condition, the HA subjects were superior.

The absence of experimental control over the learning materials and process may be a limiting factor in this study. Wide variation would be expected among the subjects as to the notes and texts used in studying for the examination, as well as for the time spent in studying for the examination.

These difficulties were substantially overcome in a study carried out by Caron (1963). He presented high school students with a 1700 word passage (consisting of an explanation of Atkinson's motive-expectancy-incentive model) to be studied in the experimental situation and, following the study period, obtained measures of rote learning and comprehension. The rote learning questions involved the reproduction of formulae and the definition of symbols contained in the passage, while the comprehension questions required the subjects to apply principles concerning risk preference that were presented in the passage. One half of his subjects studied the passage and were tested under examination conditions while the other half did so under conditions designed to induce curiosity. The condition was established by informing the subjects that the purpose of studying the passage was to enable them to interpret their own personality profiles which had been obtained in a previous testing session. For the rote learning task, there were no differences between HA and LA subjects in either treatment condition. For the comprehension task, there was no difference between HA and LA subjects in the curiosity condition. In the examination condition, however, LA subjects were superior to HA subjects. Caron (1963, p. 537) interpreted these findings as supporting the conclusion ". . . that the performance of anxious subjects on 'simple' tasks does not deteriorate under stress . . . whereas on 'complex' tasks their output suffers markedly."

While Caron's study contains many attractive features, a problem in interpreting some of his results arises because of the shortness of his measuring instruments. Only six rote learning questions and four comprehension questions were used (personal communication) and this may have operated to reduce reliability and, through this, the possibility of obtaining

significant differences between the LA and HA subjects. With respect to the rote learning task, a significant difference in favour of the LA subjects might well have been expected in the examination condition. The subjects were given only fifteen minutes to study the 1700 word passage so that learning that took place might be expected to be rather unstable and unorganized, resulting in considerable response competition in the performance situation. As has already been noted, in these circumstances anxiety may be expected to disrupt performance.

In the present study, the influence of anxiety on the performance of typical classroom tasks was again studied. As in Caron's investigation, the subjects were required to study a prose passage in the experimental situation and were then tested on several performance measures. In the present investigation, the measures obtained were of factual learning and reasoning and by increasing the number of questions asked, an attempt was made to ensure that a satisfactory level of reliability was reached for each measure. On the basis of scores on the High School Form of the Test Anxiety Scale, groups of LA, MA (moderately anxious) and HA high school students were obtained who completed the performance tasks in conditions of either high or low ego-involvement.

Hypotheses

Anxiety is conceived of as a hypothetical construct mediating between certain situational stimuli and various specifiable responses. The stimulus situation which evokes the anxiety reaction is assumed to be such that the individual anticipates a strong threat to his self-esteem. In classroom test situations, the anticipated threat to self-esteem is, most often, failure on the test.

In learning and performance situations, it is the view of Sarason and his colleagues (Mandler and Sarason, 1952; Sarason et al., 1960), that anxiety acts as a cue to elicit both responses that are relevant to the learning or performance task, and responses which are irrelevant. Task-relevant responses are observed in an increase in effort, concentration, and in procedural strategies previously found to facilitate learning and reduce anxiety. Task-irrelevant responses may be observed in the intrusion of thoughts concerning the consequences of failure, of self-depreciating ruminations and by ego-defensive avoidant responses designed to protect the individual from loss of self-esteem. These task-irrelevant responses compete with responses relevant to the task and typically have an interfering effect on learning and performance.

The extent to which interference to performance is caused by anxiety will depend upon level of ego-involvement and task complexity. When ego-involvement is low and performance is not perceived as having important ego-related consequences, little anxiety and few associated task-irrelevant responses will be elicited. In such a situation, therefore, performance for all individuals would be expected to be relatively free of the influence of anxiety. As ego-involvement increases, however, so will the tendency to react with anxiety increase and with this the tendency for interfering task-irrelevant responses to be elicited. When ego-involvement is high, individuals reacting with high levels of anxiety will respond with many more task-irrelevant responses than individuals who react to the same conditions with lower levels of anxiety. When the task is complex requiring concentration and careful processing of information, the intrusion of these task-irrelevant responses would be expected greatly to disrupt performance,

so that level of anxiety would be inversely related to performance.

In the present study, the complexity of both performance tasks was such that anxiety, when elicited, was expected to have a debilitating effect on performance. On the factual learning task, the intrusion of task-irrelevant responses was expected to interfere with both the learning and the recall of the material studied. Because of the limited exposure to the study passage, overlearning would be unlikely so that what was learned would be relatively unstable and unorganized and, as such, highly susceptible to interference resulting from anxiety. Even greater interference was expected on the reasoning task. The presence of task-irrelevant responses was expected to have a particularly disruptive effect on the application of the complex cognitive processes required for performance on this task as generalizations were made, inferences drawn and hypotheses formulated and tested.

On the basis of these considerations two hypotheses were examined:

Hypothesis 1. In low ego-involvement conditions, anxiety has no influence on performance. With both tasks, there will be no difference in the performance of LA, MA and HA groups of subjects.

Hypothesis 2. In high ego-involvement conditions, anxiety acts to disrupt performance in complex tasks. In performing both tasks, LA subjects will be superior to MA subjects and MA subjects will be superior to HA subjects.

Differences in performance for the various anxiety groups were also expected under the two ego-involvement conditions. For the factual learning task, ego-involvement was expected to facilitate the performance of LA and MA subjects. For these subjects the enhancing effects of the increased motivation induced by the high ego-involvement instructions were expected to outweigh any negative effects due to the intrusion of task-irrelevant responses associated with anxiety. Thus it was expected that their performance would be superior in the high ego-involvement condition. For the HA subjects, however, the facilitating effects of the increased motivation were expected to be completely counteracted by the interfering effects of anxiety.

With the more complex reasoning task, the interfering effects of anxiety were expected to be greater than for the factual learning task. Because of this, only the performance of LA subjects was expected to be superior in the high ego-involvement condition. For MA subjects similar levels of performance were expected for the two ego-involvement conditions. For HA subjects the interfering effects of anxiety in the high ego-involvement condition were expected to be substantially greater than any facilitating effects that might occur, so that their performance was predicted to be superior in the low ego-involvement condition.

On the basis of these expectations, two further hypotheses, concerned with difference in performance under the two ego-involvement conditions, were examined.

Hypothesis 3. With the factual learning task, the performance of the LA and MA subjects will be superior when ego-involvement is high. However, HA subjects are expected to perform no better when ego-involve-

ment is high than when it is low.

Hypothesis 4. With the reasoning task, the performance of LA subjects will be superior when ego-involvement is high, the performance of MA subjects will be similar in the two conditions of ego-involvement and the performance of HA subjects will be superior when ego-involvement is low.

Method

The subjects of the study were 173 sixth form male high school students attending three metropolitan boys' high schools in Sydney.

The content of the study passage consisted of a description of life among the Trobriand Islanders of the South Pacific. (1) This content appeared to be particularly suitable, since it was closely related to content typically taught at the high school level and yet there was little chance of the subjects having had any prior experience with it. To control the difficulty level of the vocabulary used in the passage, only words from the Thorndike-Lorge lists (1944) which occur in reading materials with a frequency of six or more times per million words were included. Thorndike and Lorge state that words appearing with this frequency are suitable for use with students in 3rd form and above. The passage contained 1332 words and one illustration, and filled almost six quarto pages of typescript.

Two performance tests were constructed. One measure, the factual learning measure, consisted of 20 multiple-choice questions for which the correct answer was explicitly stated in the study passage. The second measure, the reasoning measure, contained 12 multiple-choice questions for which the correct answer was not explicitly stated in the study passage. In answering these questions the subject was required to make deductions, and to draw inferences and implications from the given information.

Three weeks prior to the test administration, the High School Form of the Test Anxiety Scale (Mandler and Cowen, 1958), specially adapted for Australian conditions, was administered. A split-half reliability coefficient of 86 was obtained for this measure. Subjects scoring in the lower, middle and upper thirds of the anxiety distribution were designated as LA, MA and HA respectively. For each level of anxiety, the subjects were divided into two groups by use of a table of random numbers, one group being allocated randomly to the high ego-involvement condition and the other to the low ego-involvement condition.

To establish conditions of high ego-involvement (2), the subjects were informed that the test was one of scholastic aptitude and that their results would be made available to their headmaster. When the testing was completed, they were informed as to the actual purpose of the test. To establish conditions of low ego-involvement the subjects were informed that

(1) An earlier version of the study passage and performance measures was used in a previous study (Sinclair, 1965).

(2) The administration of the instruments in the high ego-involvement condition was carried out by the author in each school. The administration of the instruments in the low ego-involvement condition was carried out by T. Heys and W.J. Fenley whose assistance is gratefully acknowledged.

the purpose of the test was to establish whether the study passage was a good one for sixth form students or whether the questions were too easy or too difficult.

Twenty-five minutes were allowed for study of the passage. Twenty minutes were provided in which to answer the twenty factual learning questions and a further twenty minutes were provided in which to answer the twelve reasoning questions. These time limits were sufficient to enable all subjects to complete both tests. So that performance on the reasoning measure would not be influenced by the subjects' ability to recall information from the passage necessary for answering the questions asked, they were instructed that they could use the study passage in answering these questions.

Results

The design of the study was a 2 x 3 factorial, involving 2 levels of ego-involvement (high and low) and 3 levels of anxiety (high, moderate and low). This design was used for each of the two performance measures (factual learning and reasoning) with unequal numbers of subjects in each cell.

For the factual learning measure, the means of scores of the different anxiety groups are presented in Table 1. A reliability coefficient (K.R.20) of ·59 was obtained for this measure.

TABLE 1
Mean Factual Learning Scores for LA, MA and HA Groups of Subjects in Two Conditions of Ego-involvement

Anxiety Level	Low Ego-involvement			High Ego-involvement		
	N	$\bar{X}$	sd	N	$\bar{X}$	sd
LA	28	13·82	2·20	31	16·16	1·81
MA	28	14·32	1·94	29	14·62	2·58
HA	24	13·71	2·74	33	14·03	2·26

TABLE 2
Summary of the Analysis of Variance for the Factual Learning Measure

Source	Sum of Squares	df	Mean Square	F
Ego-involvement	41·73	1	41·73	8·17**
Anxiety	36·02	2	18·01	3·53*
Interaction ..	39·20	2	19·60	3·84*
Error	853·16	167	5·11	

** $p < ·01$.
* $p < ·05$.

A summary of the results of the analysis of variance carried out on these data (Winer, 1962, pp. 241-244) is presented in Table 2. Both main effects and the interaction were found to be significant. When individual group mean scores were examined by the Newman-Keuls procedure, it was observed that the performance of the LA group in the high ego-involvement condition had largely accounted for the significant results. As predicted, there were no significant differences found between the anxiety groups in the condition of low ego-involvement. In the high ego-involvement condition, as predicted, the performance of the LA subjects was superior to that of MA and HA subjects. The expected significant difference between the MA and HA groups did not emerge. Finally, again as hypothesized, the performance of the LA group in high ego-involvement conditions was superior to that of the LA group in low ego-involvement conditions while for the HA groups performance was similar in these two conditions. The expected superiority of the MA group in the high ego-involvement condition was not found.

TABLE 3

Mean Reasoning Scores for HA, MA and LA Groups of Subjects in Two Conditions of Ego-involvement

Anxiety Level	Low Ego-involvement			High Ego-involvement		
	N	$\bar{X}$	sd	N	$\bar{X}$	sd
LA	28	8·00	1·89	31	8·48	2·06
MA	28	7·36	2·08	29	8·38	1·82
HA	24	6·88	2·58	33	7·76	2·05

TABLE 4

Summary of the Analysis of Variance for the Reasoning Measure

	Sum of Squares	df	Mean Square	F
Ego-involvement	27·16	1	27·16	6·28*
Anxiety	24·77	2	12·39	2·87
Interaction	2·23	2	1·11	
Error	721·68	167	4·32	–

* $p < .05$.

In sum, the hypothesized relationships for the LA and HA groups in the two conditions of ego-involvement were all confirmed. Those for the MA group were not confirmed, the performance of that group being no different from that of the HA group.

For the reasoning measure, the mean scores of the different anxiety groups are presented in Table 3. A reliability coefficient (K.R.20) of ·68 was obtained for this measure.

A summary of the results of the analysis of variance carried out on these data is presented in Table 4. In this analysis only the mean square for level of ego-involvement was significant, indicating a general superiority in the high ego-involvement conditions. When pairs of means were analysed, again using the Newman-Keuls procedure, it was found that there were no differences between the anxiety groups in either ego-involvement condition. This was predicted for the low ego-involvement condition but for the high ego-involvement condition an inverse relationship between level of anxiety and performance had been predicted. All anxiety groups performed better in the high ego-involvement condition (although in no case did the difference reach an acceptable level of significance). This was predicted for the LA subjects but not for the MA and HA groups. In fact, for HA subjects superior performance had been predicted for the low ego-involvement condition.

Discussion

With respect to the factual learning task, the results obtained confirmed, in large measure, the hypotheses that were developed for testing. In test-like conditions, anxiety was observed to debilitate performance on that task. With respect to the reasoning task, however, few predicted relationships were supported. Despite the complexity of the task, anxiety did not appear to influence performance in the test-like condition. A possible reason for this latter result is to be found in the manner in which the reasoning test was administered. So that all subjects would have approximately equal access to the factual information upon which the reasoning items depended, the subjects were allowed to consult the study passage while answering the questions. This would make the reasoning task rather comparable to an open-book examination in which the student is able to consult certain reference material on answering the question asked. This procedure, by providing a memory-support (Sieber, 1969) in the performance situation, may well have had a reassuring, anxiety-reducing effect on the HA subjects so that interference to performance due to anxiety may have been minimal.

The results obtained provide a number of conclusions that bear directly on classroom practice and on the different theories that have been developed to explain the influence of anxiety on learning and performance. With respect to the factual learning task, the results support the conclusion that anxiety operates to debilitate performance when a complex task is to be performed in test-like conditions. This conclusion suggests that in important examinations, the HA student will be at a considerable disadvantage. When competing with other students for scholarships, university entrance, school prizes, employment opportunities or simply place in class, anxiety will act to interfere with and reduce the level of his performance.

The results also support the conclusion that while instructions designed to increase level of ego-involvement will raise the level of performance of LA students, it will not do so for MA and HA students. Sarason's theory suggests that for the MA and HA student, the positive motivational benefits deriving from the ego-involving instructions are cancelled out by the operation of task-irrelevant responses which are also

elicited.

This conclusion suggests that the widely adopted practice in education of attempting to motivate students by placing strong emphasis upon the importance of examinations and the need to do well and avoid failure will be of value only to low test anxious students. In the present study with respect to the performance of moderately and high test anxious students on the factual learning task, little was achieved by increasing level of ego-involvement and, through this, anxiety. In fact, it may be that this emphasis, from a long term view, will have quite harmful effects. Since, at high levels, anxiety is such an unpleasant and exhausting experience, this emphasis may serve to engender a strong dislike of school which may eventually lead the student to drop out of school prematurely. Some support for this possibility is provided by Spielberger (1962) who observed, in one study, that HA college students had a higher drop out rate than LA students of comparable ability.

In addition to the implications provided for education practice, the results of the present study also provide implications for theory. The conclusion that in a test-like situation, anxiety will interfere with performance on a complex task is, as we have seen, consistent with the viewpoint of Sarason and his colleagues (Mandler and Sarason, 1952; Sarason et al., 1960). It is also, however, consistent with the Spence-Taylor theory, although in this theory it is the drive function of anxiety that is emphasised rather than the cue function. Spence and Taylor (Spence and Spence, 1966), conceive of anxiety as a drive which combines multiplicatively with the habit strengths of responses present in the individual's response hierarchy. When the desired response is not clearly dominant in the response hierarchy, as tends to be the case in complex performance situations, increase in drive (anxiety) serves to heighten competition among potential responses and in so doing disrupts performance.

The conclusion reached that increase in level of ego-involvement (stress) serves to raise the performance of LA individuals but not MA and HA individuals is, again, consistent with Sarason's theory. This conclusion, however, is not easily accounted for by the Spence-Taylor theory. Although, in the most recent statement of their position (Spence and Spence, 1966), they give passing reference to the question of situational factors (such as ego-involving instructions) that serve to elicit anxiety, they have not considered this question in detail, nor attempted to manipulate such factors in their research studies.

A number of directions for future research are suggested by the results of the present study. In this study the subjects used were male and of above-average ability. There is a need, then, for research to be carried out to determine if the conclusions reached in this study also hold for females and students of average and below-average ability. It is important, too, that ways be found to control the interfering effects of anxiety in the classroom. In particular, ways need to be found by which the HA student may be challenged but his anxiety kept within non-debilitating limits. One suggestion that arises from the present study is the possibility of using open-book examinations where reasoning is the major objective of assessment. Being able to consult appropriate reference material in the examination situation reduces the strain of having to remember and recall large bodies of information and in so doing may serve to reduce anxiety and the interference to reasoning that results. Sieber (1969) in an important recent article, provides further experimental evidence that

the provision of memory supports will be a particular aid to HA students in counteracting the interfering effects of anxiety. In that article she also suggests a number of other ways by which the HA student may be helped to perform more effectively. In particular she discusses the benefits that may be derived from instruction in the use of verbal encoding skills, diagrams, mnemonic devices, notational systems and outlining systems for organizing general ideas prior to the development of detail. There is a need for these suggestions to be followed up in classroom-oriented research.

Bibliography

Caron, A.J. "Curiosity, Achievement, and Avoidant Motivation as determinants of Epistemic Behaviour" J. abnorm. soc. Psychol., 67, 1963, 535-49.

Deese, J., Lazarus, R.S., and Keenan, J. "Anxiety, Anxiety Reduction and Stress in Learning" J.exp. Psychol., 46, 1953, 55-60.

Feshbach, S, and Loeb, A. "A Further Experimental Study of a Response-interference versus a Drive-facilitation Theory of the Effect of Anxiety upon Learning" J. Personal., 27, 1959, 497-506.

Harleston, B.W. "Test Anxiety and Performance in Problem solving Situations." J. Personal, 30, 1962, 557-573.

Hebb, D.O. "Drive and the C.N.S. (Conceptual Nervous System)." Psychol. Rev., 62, 1955, 243-254.

Kalish, H.I., Garmezy, N., Rodnick, E.H., and Bleke, R.C. "The Effects of Anxiety and Experimentally induced Stress on Verbal Learning." J.gen. Psychol., 59, 1958, 87-95.

Longnecker, E.D. "Perceptual Recognition as a Function of Anxiety, Motivation, and the Testing Situation." J. abnorm. soc Pyshchol., 64, 1962, 215-221.

Lucas, J.D. "The interactive Effects of Anxiety, Failure, and Intraserial Duplication." Amer. J. Psychol., 65, 1952, 59-66.

Malmo, R.B. "Activation: a Neuropsychological Dimension." Psychol. Rev., 66, 1959, 367-86.

Mandler, G., and Cowen, J.E. "Test Anxiety Questionnaires" J. consult. Psychol., 22, 1958, 228.

Mandler, G., and Sarason, S.B. "A study of Anxiety and Learning." J. abnorm. soc Psychol., 47, 1952, 166-173

Montague, E.K. "The Role of Anxiety in Serial Rote Learning" J. exp. Psychol., 45, 1953, 91-98.

Nicholson, W.M. "The Influence of Anxiety upon Learning: Interference or Drive Increment?" J. Personal., 26, 1958, 303-319.

Paul, G.L., and Eriksen, C.W. "Effects of Test Anxiety on 'Real-Life' Examinations." J. Personal., 32, 1964, 480-494.

Sarason, I.G. "The Effect of Anxiety, Motivational Instructions and Failure on Serial Learning," J. exp. Psychol., 51, 1956, 253-259.

Sarason, I.G. "Effect of Anxiety and Two Kinds of Motivating Instructions on Verbal Learning" J. abnorm. soc. Psychol.54, 1957a, 166-171.

Sarason, I.G. "The Effect of Anxiety and Two Kinds of Failure on Serial Learning", J. Personal., 25, 1957b, 383-392

Sarason, S.B., Davidson, K.S., Lighthall, F.F. Waite, R.R., and Ruebush, B.K. Anxiety in Elementary School Children, New York, John Wiley, 1960.

Sieber, J.E. "A Paradigm for Experimental Modification of the Effects of Test Anxiety on Cognitive Processes." Amer, Educ. Res.J. 6, 1969, 46-61.

Sinclair, K.E. The influence of Anxiety and Level of Ego Involvement on Classroom Learning and Performance. Unpublished Ph.D.

dissertation, University of Illinois, 1965.
Spence, J.T., and Spence, K.W. "The Motivational Components of Manifest Anxiety: Drive and Drive Stimuli" in C.D. Spielberger (Ed.), Anxiety and Behaviour. New York: Academic Press, 1966, 191-326.
Spence, K.W. and Taylor, J.A. "Anxiety and Strength of the UCS as Determinants of the Amount of Eyelid Conditioning," J. exp. Psychol., 42, 1951, 183-188.
Spielberger, C.D. "The Effects of Manifest Anxiety on the Academic Achievement of College Students." Ment. Hyg., 46, 1962, 420-426.
Standish, R.R., and Chamption, R.A. "Task Difficulty and Drive in Verbal Learning". J. exp Psychol. 59, 1960, 361-365.
Taylor, J.A. "The Relation of Anxiety to the Conditioned Eyelid Response". J. exp. Psychol., 41, 1951, 81-92.
Taylor, J.A. and Spence, K.W. "The Relationship of Anxiety Level to Performance in Serial Learning. "J. exp. Psychol., 44, 1952, 61-64.
Thorndike, E.L., and Lorge, I. The Teacher's Word Book of 30,000 Words. New York: Bureau of Publications, Teachers College Columbia University, 1944.
Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.
Wrightsman, L. S. "The effects of Anxiety, Achievement Motivation, and Task Importance upon Performance on an Intelligence Test." J. educ. Psychol., 53, 1962, 150-156.

A SELECTION OF CURRENT RESEARCH PROJECTS ON EXAMINATIONS*

EDUCATIONAL PLANNING

Name of Institution	-	School of Education, University of Zambia.
Title of Project	-	REPORT OF THE COMMITTEE ON POST-CAMBRIDGE SCHOOL CERTIFICATE STUDIES AND UNIVERSITY ENTRANCE.
Name of Persons and Designation	-	A Committee of the Board of Studies, School of Education, comprising Trevor Coombe, Senior Lecturer in Education, and David Stannard, Lecturer in Education.
Main Aim of Project	-	The terms of reference of the Committee were as follows: 'to assemble evidence from the school system and draw conclusions upon the desirability of re-introducing post-Cambridge Overseas School Certificate instruction in Zambian secondary schools as a preparation for entrance to the University of Zambia'.
Description	-	The body of the report comprises a discussion of the probable condition of several salient aspects of the Zambian school system up to 1975. The discussion deals with the methodology and results of projections in the following areas: secondary enrolment and candidates for the Cambridge Overseas School Certificate examination; performance in the Cambridge examination; first-year full-time University of Zambia enrolment direct from school (degree candidates only); and the secondary teaching force. The report also contains an introduction and a conclusion, which recommends against the revival of large-scale post COSC studies in the Zambian secondary system, at least until 1975. The report, which is mimeographed, consists of 36 pages of text and a 35-page appendix of tables and figures.
Date of Commencement	-	March 26, 1969.
Expected Date of Completion	-	September 19, 1969.
Source of Finance	-	None required
Method and Anticipated Date of Publication	-	It is possible that the report will be published in more permanent form along with a selection of official and unofficial papers dealing with the question of the level of entrance to the University of Zambia.

*Reprinted from Education in the Developing Countries of the Commonwealth; Abstracts of Current Research, 1969, Commonwealth Secretariat, London, 1970.

METHODS OF ENQUIRY

Research methods; evaluation procedures

Name of Institution	-	School of Education, University of the South Pacific, Suva, Fiji.
Title of Project	-	EDUCATIONAL TESTING.
Name of Person and Designation	-	R.C. Honeybone - Professor of Education.
Main Aim of Project	-	The School of Education has been given the task of looking into the possibility of developing improved measures for educational assessment and selection in Fiji.
Description	-	Of prime importance is the development of a test of general ability. The normal test construction problems of developing countries are increased by the necessity to ensure that results of test variance in the two major racial groups (Melanesian and Indian) are not too dissimilar. Initial results indicate that further problems may also be added by maturational and developmental differences, between urban and rural groups. We have started one research project in the field of Educational Testing and will be setting up an assessment unit at the beginning of 1970.

Name of Institution	-	Monash University, Australia.
Title of Project	-	CLASSIFICATION AND ORDERING SKILLS IN YOUNG CHILDREN.
Name of Person and Designation	-	Miss Mary Creighton Nixon - Senior Lecturer.
Main Aim of Project	-	(i) to establish whether test material used with Australian children could be used with indigenous children in the Territory of Papua - New Guinea; (ii) if so, to obtain data for comparison purposes; (iii) to apply findings to methods of teaching subjects in which the use of 'logical systems' is necessary.
Description	-	The first aim was investigated by testing standard III Tolai children in the Gazelle Peninsula in New Britain on a cross classification task. They were able to carry this out at about the level of 7 years old white Australian children. The work has not proceeded further.
Date of Commencement	-	February, 1968.

Name of Institute	-	Department of Education, University of Hong Kong.
Title of Project	-	MODERN EXAMINING.
Name of Person and Designation	-	Group of Local Examiners.
Description	-	A summary and analytical and critical account of examining techniques, with examples and specially applied to Hong Kong's school examination system.
Date of Commencement	-	September, 1969.
Expected Date of Completion	-	June, 1971.
Source of Finance	-	Department of Education, University of Hong Kong.

Name of Institution	-	National Council of Educational Research and Training, India.
Title of Project	-	CO-OPERATIVE TEST DEVELOPMENT PROJECT - INTELLIGENCE TEST.
Name of Persons and	-	Dr. (Mrs.) Uma Sinha - Reader, Dr. K. Kulandaivel - Principal, Sri Ramakrishna Mission Vidayalaya Teachers College, Coimbatore, Dr. (Mrs.) K. Pramanik - Principal, Institute of Education for Women, Calcutta, Dr. N.P. Pillai - Professor of Education, Kerala University, Kerala.
Main Aim of Project		The main aim of the project is to develop four group tests of intelligence each with a parallel form and partial overlap for the ages 7+ to 9+, 9+ to 11+, 13+ to 16+ and 11+ to 13+.
Description	-	Four tests each with a parallel form from the above four age levels are being developed. There is an overlap of age between the two consecutive age levels. This is to develop one scale of intelligence from 7+ to 16+ by scaling the adjacent level tests so that prediction of the scores on later tests may be made on the basis of scores on earlier tests. The tests are for Verbal Educability Factor. They will thus be verbal and quantitative items so that they give two separate scores for V and N in addition to one general score. The tests are being developed in three languages, i.e. Malayalam, Tamil and Bengali. Later on the tests will be adopted to other Indian languages so that we have intelligence tests at national level.

Date of Commencement	-	November, 1966.
Expected Date of Completion	-	March, 1971.
Source of Finance	-	Funds of N.C.E.R.T.
Method and Anticipated Date of Publication	-	The report will be published by NCERT sometime during 1971.

Name of Institution	-	National Council of Educational Research and Training, India.
Title of Project	-	CO-OPERATIVE TEST DEVELOPMENT PROJECT - INTEREST INVENTORY.
Name of Persons and Designation	-	Dr. (Mrs.) Uma Sinha - Reader, Professor B. Krishnan - Professor of Psychology, Mysore University. Professor M.Y. Bhide - Vice-Principal, Sadhna School of Educational Research and Training, Bombay. Dr. B. De - Director, Bureau of Educational and Vocational Guidance, Patna, Mr. V.G. Jhingran - Principal D.S. College, Aligarh.
Main Aim of Project	-	To develop a vocational Interest Inventory with scales for a number of occupations.
Description	-	The vocational Interest Inventory which is being developed is for the age group 14-25. The Inventory is not on any existing model. New types of items are being tried. 13 occupational areas, namely Social Services, Economics, Secretarial, Administrative, Technical, Outdoor Physical, Scientific, Linguistic Literary, Arts and Entertainment, Protective, Educational, Medical and Religious have been taken into account while developing items. Scales for only eight to ten such occupations which are likely to absorb relatively larger numbers of individuals will be developed.
Date of Commencement	-	November, 1966.
Expected Date of Completion	-	March, 1971.
Source of Finance	-	N.C.E.R.T. funds.
Method and Anticipated	-	The preliminary report will be published by the NCERT some time in 1971.

Name of Institution	-	National Council of Educational Research Training, India.
Title of Project	-	DIFFERENTIAL APTITUDE TESTS.

Name of Persons and Designation	-	Mrs. S. Shukla - Reader, P.M. Patel - Lecturer J.P. Mittal - Technical Assistant.
Main Aim of Project	-	To develop a Differential Aptitude Test Battery for the Hindi-speaking region for allocation of students to various courses being offered at the higher secondary stage.
Description	-	It is planned to improve upon and standardize a series of tests - already developed initially in one of the States in India for a small geographical region - in verbal, numerical and spatial abilities for students of age group 13 to 14. The initial spade work has already been done. This battery would be standardized for the entire Hindi-speaking region. Differential validity would be made available to the users.
Date of Commencement	-	March, 1969.
Expected Date of Completion	-	March, 1972.
Source of Finance	-	NCERT funds.
Method and Anticipated Date of Publication	-	1973.

Name of Institution	-	Department of Educational Psychology, Makerere University College, Uganda.
Title of Project	-	THE INTELLECT OF EAST AFRICAN STUDENTS.
Name of Person and Designation	-	Hamed El-Abd - Professor of Educational Psychology and Head of Department.
Description	-	Investigation based on the assumption that the structure of intellect is independent of race and religion, but dependent on education and experience, was conducted using an East African sample from undergraduates and high school certificate students.
Date of Commencement	-	November, 1965.
Expected Date of Completion	-	Research continuing as long as funds are available.
Source of Finance	-	(a) The Nuffield Foundation, through authorization by the Institute of Education, Kampala. (b) Makerere University College Research Grant.
Method and Anticipated Date of Publication	-	Mimeographed reports are produced at present by the Educational Psychology Department. Funds are anticipated in the future to make possible printed booklets of the research material.

Name of Institution	-	The West African Examinations Council, Lagos, Nigeria.
Title of Project	-	FEASIBILITY OF THE MRC 1501 ANSWER CARD.
Name of Person and Designation	-	Dr. D.J. Fitch - Technical Adviser.
Main Aim of Project	-	To investigate the feasibility of using the 1501 in West Africa.
Description	-	To investigate the feasibility of using the 1501 in West Africa, the Research Evaluation, and Statistical Analysis Section has designed a study. The Measurement Research Corporation has provided 1501 answer cards of a standard design. These were used in the States of Northern Nigeria to administer a form of the CEE. The same test was administered in other schools using the 1230 sheet. The study will allow us to compute test-retest and parallel forms reliability using each document and thereby make an important comparison.
Date of Commencement	-	July, 1969.
Expected Date of Completion	-	December, 1969.
Source of Finance	-	W.A.E.C. and USAID.
Method and Anticipated Date of Publication	-	TEDRO Report December, 1969.

Name of Institution	-	West African Examinations Council, Lagos, Nigeria.
Title of Project	-	FORCED CHOICE RATING FORMS.
Name of Persons and Designation	-	Dr. D.J. Fitch - Technical Adviser, Dr. S.A. Akeju - Principal Research Officer.
Main Aim Project	-	To develop a forced choice rating form for nurses.
Description	-	With the usual rating procedure one is asked to evaluate the ratee in terms of his position on several trait scales. Typically such scales have five points. One might be asked to say if the ratee were (a) very hard working (b) moderately hard working (c) average (d) moderately lazy or (e) very lazy. The main trouble with such a rating form is that he can make a ratee "look good" or "look bad". He typically is motivated to make a ratee "look good" but not too good. He reasons that other raters who are rating other people have similar motives and he wants his ratees to generally compare favourably with those of other raters but he is not sure how the other

		will rate so he does not know how to rate his own ratees. What would be desirable would be a form which would allow an accurate and honest description of behaviour that would not require or allow the rater to directly make good-bad decisions. The forced choiced rating form meets this need.
Date of Commencement	-	Within the next year.
Expected Date of Completion	-	Not known.
Source of Finance	-	W.A.E.C. and USAID.
Method and Anticipated Date of Publication	-	TEDRO REPORTS.

Name of Institution	-	The East African Examinations Council, Lagos, Nigeria.
Title of Project	-	GAMBIA COMMON ENTRANCE EXAMINATION.
Name of Person and Designation	-	Dr. D.J. Fitch - Technical Adviser
Main Aim of Project	-	To provide information on the extent to which the tests developed in Nigeria are appropriate for use in The Gambia.
Description	-	The same CEE test has been used in the Gambia and Nigeria. Headmasters in the Gambia have asked as if we can provide them with any information on the extent to which these tests developed in Nigeria are appropriate for use in the Gambia. We plan to undertake reliability studies using 1969 data from the two countries separately. A validity study is also being conducted in the two countries.
Date of Commencement	-	August, 1969.
Expected Date of Completion	-	December, 1969.
Source of Finance	-	W.A.E.C. and USAID.
Method and Anticipated Date of Publication	-	TEDRO Report - December, 1969.

Name of Institution	-	The West African Examinations Council, Lagos, Nigeria.
Title of Project	-	VALIDATION OF ACHIEVEMENT TESTS.
Name of Person and Designation	-	Dr. D.J. Fitch - Technical Adviser.
Main Aim of Project	-	Validation studies to increase the accuracy and efficiency by which achievement can be measured.

Description	-	A review of the Godschalk, Swineford and Coffman study undertaken at ETS and published in 1966 will be carried out with subject matter panels. From this will be developed a highly reliable and valid a achievement criterion for their subject and a large experimental battery of short traditional and objective papers. The criterion measure would be given to a sample of a few hundred students along with the experimental battery. Papers of the battery would be validated against the criterion and a subset of papers predicting the criterion would be selected. The possibilities of increasing the accuracy and efficiency by which achievement can be measured are very large by means of such research.
Date of Commencement	-	Hopefully in the near future.
Expected Date of Completion	-	Not known.
Source of Finance	-	W.A.E.C . and UNSAID.
Method and Anticipated Date of Publication	-	TEDRO Reports.

Name of Institution	-	The West African Examinations Council, Lagos, Nigeria.
Title of Project	-	STANDARD AND SIDE BY SIDE ANSWER SHEETS.
Name of Person and Designation	-	Dr. D.J. Fitch - Technical Adviser
Main Aim of Project	-	Designing of standard answer sheets that could be used with most or all of our tests.
Description	-	On large testing programms it is highly desirable to have answer sheets computer printed with the candidate's name and his index number grid printed in a machine readable form. When such answer sheets are returned for scoring the candidate's number along with his scores are punched into a card. If the candidate himself had marked the index number grid there would be a chance that a mistake would be made. This could mean losing the candidate's scores or, worse still, assigning one candidate's scores to another candidate. A difficulty with such computer printing is that it is quite time consuming. One can expect to print less than 1,000 per hour. Such printing can be done at twice the speed if answer sheets are printed side by side.
Date of Commencement	-	February, 1968.
Expected Date of Completion	-	Studies expected to continue indefinitely.

Source of Finance	-	W.A.E.C. and USAID.
Method and Anticipated Date of Publication	-	No reports planned.

Name of Institution	-	Institute of Education, University of Ibadan, Nigeria.
Title of Project	-	PREPARATION AND STANDARDISATION OF READING TESTS FOR NIGERIAN SCHOOL CHILDREN.
Name of Person and Designation	-	Dr. John O.O. Abiri - Lecturer of Education.
Main Aim of Project	-	To prepare a set of standardized Reading Tests (initially in English, but possibly in at least one of the major Nigerian languages later) which may be used for grading, selection and comparative assessment of pupil's achievements in reading.
Description	-	Writing, trial runs, and standardization of test materials.
Date of Commencement	-	As soon as funds are available.
Expected Date of Completion	-	About two years from date of commencement.
Source of Finance	-	None yet.

Name of Institution	-	Department of Institute of Education, University of Ibadan.
Title of Project	-	ADAPTING OBJECTIVE TESTS FOR LOCAL USE.
Name of Person and Designation	-	Dr. E.A. Yoloye - Lecturer in Education.

STUDENTS' BACKGROUND AND CHARACTERISTICS

Name of Institution	-	Department of Educational Psychology, Makerere University College, Uganda.
Title of Project	-	CONSTRUCTION AND APPLICATION IN EAST AFRICAN SCHOOLS OF THE GENERAL VERBAL ABILITY TESTS.
Name of Person and Designation	-	Hamed El-Abd - Professor of Educational Psychology and Head of Department.
Description	-	These tests are designed for use in the upper primary classes and junior secondary grades of East African schools. The tests could also be applied to children of similar education and cultural backgrounds.
Date of Commencement	-	November, 1965.
Expected Date of Completion	-	Research continuing as long as funds are available.
Source of Finance	-	(a) The Nuffield Foundation, through authorization by the Institute of Education, Kampala. (b) Makerere University College Research Grant.
Method and Anticipated Date of Publication	-	Mimeographed reports are produced at present by the Educational Psychology Department. Funds are anticipated in the future to make possible printed booklets of the research material

SELECTION AND PERFORMANCE OF STUDENTS

Name of Institution	-	Department of Educational Psychology, Makerere University College.
Title of Project	-	THE LULE NON-VERBAL INTELLIGENCE TESTS.
Name of Person and Designation	-	Hamed El-Abd - Professor of Educational Psychology and Head of Department.
Description	-	The original tests were designed by Mr. Lule, the present Principal of Makerere College, in 1952 for use in multi-racial schools in Africa. The new edition of the tests was revised during 1966 and 1967.
Date of Commencement	-	November, 1965.
Expected Date of Completion	-	Research continuing as long as funds are available.
Source of Finance	-	(a) The Nuffield Foundation, through authorization by the Institute of Education, Kampala. (b) Makerere University College Research Grant.
Method and Anticipated Date of Publication	-	Mimeographed reports are produced at present by the Educational Psychology Department. Funds are anticipated in the future to make possible printed booklets of the research material.

Name of Institution	-	Human Development Research Unit, University of Zambia.
Title of Project	-	ORIENTATION ERRORS IN COPYING KOH'S BLOCKS DESIGNS.
Name of Person and Designation	-	Dr. Robert Serpell - Lecturer in Psychology
Main Aim of Project	-	Analysis of the causes of orientation errors in copying tasks.
Description	-	Following the 'response organisation' hypothesis proposed in a previous report (Serpell, R. (1969) Cross-cultural differences in the difficulty of copying orientation: a response organisation hypothesis. HDRU Reports, 12, Lusaka: University of Zambia - mimeo.) a study has been initiated of the influence of sequential response organisation on the orientation of Koh's Blocks reproductions. Responses of primary school children to the first 5 items of the regular test have been recorded by an overhead cine-camera. Further trials are now being undertaken with selected designs for which the hypothesis generates exact predictions.

Date of Commencement	-	May, 1969.
Expected Date of Completion	-	October, 1970.
Source of Finance	-	British Government, Ministry of Overseas Development Grant to University of Zambia for Human Development Research Unit.
Method and Anticipated Date of Publication	-	HDRU Report in 1970, subsequently in international journal.

OTHER EDUCATIONAL PUBLICATIONS BY THE COMMONWEALTH SECRETARIAT

1) Education and Training of Technicians

Report of a Conference held at Huddersfield, England, 1966. (pp.326. Price 22s.6d. net).

Commonwealth experts in technical education discuss a wide range of important issues: manpower planning; types, content and organisation of technician courses; planning and equipping technical institutions; training technical teachers; selection of trainees; standards of attainment.

2) Mathematics in Commonwealth Schools

Report of a Conference in Trinidad, 1968, (pp.199. Price 20s. net).

A practical approach to the teaching of mathematics at primary and secondary level, reflecting the experience of leading mathematicians and other educationists in Commonwealth countries. Extensive bibliography.

3) Youth and Development in Africa

Report of the Commonwealth African Regional Youth Seminar, held at Nairobi, November, 1969. (pp.220. Price 20s. net).

An important new book for those concerned for the future of Africa and with the role of young people in development.

4) Education in Rural Areas

Report of a Special Conference held in Accra, March 1970. (pp. 340 approx. Price 10s. net).

5) Commonwealth Scholarship and Fellowship Plan Annual Reports

The Commonwealth Scholarship and Fellowship Plan was drawn up at the First Commonwealth Education Conference held at Oxford in 1959. The basis of the Plan is that those Governments which are able to do so should institute fellowships to be held by men and women from other Commonwealth countries at universities and other institutions of higher learning in the awarding country. At present nearly twenty countries of the Commonwealth award fellowships under the Plan while nominations are received from all the Commonwealth countries. The Annual Reports provide summary reports from the countries on the working of the Plan and useful statistical analyses. (Price 5s.)

6) Education in the Developing Countries of the Commonwealth: Abstracts of Current Research, 1969. (Price 5s. net)

Sixty universities and other institutions in the Commonwealth provide details of their current research activities on educational topics.

Published by the
COMMONWEALTH SECRETARIAT

To be purchased from the
Commonwealth Secretariat
Printing Section
Marlborough House
London, S.W.1.

S.B.N. 85092 034 5