

II. TRENDS AND DEVELOPMENTS IN EXAMINATION TECHNIQUES

A Survey of Needs and Approaches

More than one commentator on examinations has expressed the hope that perfection would never be attained. Imperfect examinations, it has been suggested, allow the worst examination failures to retain a modicum of self-respect because it is possible to blame lack of success on the imperfection of the instrument, whereas failure in perfect examinations could be interpreted only as an indication of personal inadequacy. Ten years ago one examination board secretary wrote, "That will be a sorry day for human initiative and personal aspirations on which there is announced the examination which shall be completely valid, the perfect predictor of each individual's ineluctable future."* That day still lies far in the future. Nevertheless, the development of markedly better means of testing and measurement than those available today would provide improved indicators for individuals and educators while still falling short of perfection. A realistic and reasonably accurate assessment of achievement and potential can lead to a recognition of attainable goals in life. In this way good examinations can contribute both to individual satisfaction and national development.

Tests of attainment, especially in certain prescribed areas, have reached a generally acceptable level of validity and reliability. A range of approaches, varying in their objectivity, give results at an acceptable level of accuracy. Similarly, diagnostic tests validated in particular settings have become useful aids to teaching and learning. Many other types of testing, however, are less perfect, particularly those intended to provide information about aptitude and likely future performance. Testing in domains other than the cognitive remains largely undeveloped. Tests in spoken language and practical skills often form only a small proportion of the overall examination because of the difficulty of ensuring their accuracy. The imperfection of the single examination at the conclusion of a course is well known.

To meet the criticisms, new approaches and new forms of examination are being developed. Increased objectivity, a move towards testing over a period of time, the development of batteries of associated tests of different kinds, a closer relationship between the declared objectives of the syllabus and the content and form of the examination, the reduction of the influence of chance and other extraneous factors, all have become major preoccupations of examinations research and development in recent years.

Each testing technique has its own advantages and usefulness and it is the duty of an examining body constantly to review procedures best adapted to each curricular goal. An examining body would perhaps abdicate its responsibility if it believed that its duty was only to examine, without reference to the wider context of educational change. Consideration should be given to all the requirements of the various objectives of the curriculum. The tests produced by examining bodies, therefore, might on occasion range beyond the examination syllabuses in their present form. The close link between tests and the objectives of teaching should be emphasised. Syllabuses should set out the kind of goals they expect teaching to achieve and examinations should then be tailored to these goals.

* J.A. Petch. G.C.E. and Degree, Part 2, OP.14, Manchester, Joint Matriculation Board, September 1963, p. 119.

The fact that large numbers of candidates who graduate from schools do not gain admission to higher institutions makes it necessary for educational institutions to emphasise the terminal aspects of education. Examining bodies should seek to devise a range of instruments for meeting the needs of both terminal and continuing candidates.

Some Specific Problem Areas

Problems arise from the need to improve examining instruments related to existing curricula, syllabuses and needs and also from the need to develop new examinations in association with new educational content and method. Unless constant care is exercised, educational desirability may be subordinated to administrative convenience, those aspects of a syllabus being examined which it is possible to examine rather than those which educational criteria indicate should be examined. For this reason the declared objectives of curricula may not be examined effectively. It must be recognised, however, that educational principles must be reconciled with administrative possibilities and financial resources in order to achieve the best compromise in the prevailing circumstances.

Problem areas related to existing syllabuses include the examining of spoken language and practical science as well as the inclusion of an element of choice in examinations. The problem of developing "culture fair" tests which do not discriminate against disadvantaged candidates presents particularly acute and pressing difficulties. New approaches to education, such as the large-scale introduction of pre-vocational subjects in Sri Lanka or of technical subjects in Singapore, call for the creation of new means for measurement and prediction. Perhaps a whole new philosophy of examining is needed.

Beyond the immediate problems relating to the form of examinations lies the complex area of public opinion, educational respectability and examination credibility. When attempting to introduce a radically revised syllabus, educational policy makers encounter immediate suspicion on the part of teachers, pupils and parents who view innovation as a tampering with recognised material, values and standards. The need to adjust to new teaching practices and new syllabus content is seen by teachers not infrequently as an unnecessary burden.

It is always easier to add to the curriculum than to prune it. Innovations without deletion can lead to overburdened syllabuses which mean that examinations can cover the field only patchily. This in turn may lead to pupils covering an unbalanced series of items from the overall syllabus.

The introduction of new curricula is often justified on the grounds of increased relevance to current needs and a move away from the academic curriculum. Even if it is recognised that the "academic" curriculum is in fact effective vocational training for the professional men, administrators and clerks whom employers have seen as the prime justification for the existence of schools, it remains unproved that the academic curriculum is basically responsible for the inculcation of that distaste for technical and manual occupations which has caused so much concern to governments over the last century. The solution may well lie not only in more "relevant" curricula and examinations but in a more effective integration of the educational process with national needs and national development, to the end of ensuring that those who engage in technical and manual occupations receive returns - financial and in terms of status - which enable them to enjoy a similar quality of life to those engaged in "white collar" pursuits.

Even when public opinion and the teaching profession have been won over to the acceptance of new forms and content of education, there remains the problem of devising suitable examining procedures, educationally valid and administratively possible. The difficulty of standardising traditional examination papers in which there are optional questions, for example, may be illustrated by the fact that if candidates are asked to choose four out of nine questions on each of two papers in a subject, nearly 16,000 question combinations are possible. In this case, since each candidate selects his individual set of questions to answer, it is inevitable that the choice of questions will vary among candidates, even though they are sitting what is nominally the same examination. Yet if choice is eliminated it seems necessary for the examination to cover the whole syllabus and to introduce a degree of compulsion which has been avoided in the past. The need seems to be for the creation of a revised form of examination, probably including as one element a series of short, objective-type questions.

Objective-type examinations may escape some of the difficulties associated with the subjective essay-type questions but at their present stage of development they restrict the nature of the testing possible. These examinations, like essays, cannot yet cover all aspects of all subjects currently taught on a wide scale, while little is yet known of the "backwash" effects of such examinations on the teaching-learning process in schools. In the light of current knowledge, however, it appears that both traditional and newer types of question can play a useful part in a complete examination.

The difficulties presented by large and increasing numbers of candidates, often dispersed over wide geographical areas, affect in particular the administration of oral and practical tests. That orals and practicals are difficult to validate and standardise is well known; that it is desirable to examine in this way for some purposes at some levels is generally accepted. As their cost approaches the prohibitive a point is reached at which educational and administrative interests may clash.

Large-scale examining necessarily involves the use of large numbers of teachers, with possible problems of security and competence. Such participation by teachers may expose them to severe pressures from parents, pupils and inspectorate.

Further problems arise from the use of examinations for more than one purpose. Selection tests for the few often serve as terminal examinations for the many; as predictors, such examinations are rarely efficient; as measures of attainment they record only fractional information about the candidate upon which potential employers can base their judgment. Often, too, an examination which is terminal in the context of the formal educational system also acts as the first selection test for entry into employment and may be a far from perfect instrument for this purpose.

The problems associated with examination techniques are multifarious and persistent. The search for relevance, validity, reliability and objectivity continues. Here the contribution of carefully directed research and evaluation is crucial, so that examinations may be developed to meet the needs of the candidates and the nation rather than training candidates to meet the demands of inappropriate testing procedures.

Some Examples of New Approaches

The quest for suitable examining techniques has led beyond simple improvements of existing practices into new approaches and new forms of examinations more closely tailored to the ends which they are supplied to serve.

One area in which further development could be of much value is in informal tests for classroom use. Every teacher in the course of his career will prepare, administer and mark a large number of tests. By them he makes judgments on his pupils and receives feedback on the effectiveness of his teaching. If teachers can be helped to acquire the techniques needed to build and evaluate better informal tests their value could be reinforced and their contribution to continuous assessment enhanced. In some circumstances these tests might be developed into a co-ordinated sequence of the type known as "continuous achievement monitoring". These tests might be extended to include formal assignments or project work. There is some risk, however, that such an extended coverage might over-formalise a system which many competent teachers employ already.

The value of self-evaluation and peer evaluation does not seem to have been fully appreciated and further development in this area would be useful. Care must be taken, however, that these techniques do not lead to an excessively competitive and negative reaction, self-derogatory and critical rather than positive and appreciative. As education is likely to move increasingly in the direction of informal and individual learning, self-evaluation will acquire an increasing importance so that improvements in this technique will be of much practical utility.

The constraints of university selectors on the curriculum and examination process are widely recognised. Several notable developments have been designed to meet this problem. The International Baccalaureate provides a flexible syllabus leading to an examination which is accepted by university authorities in many countries as qualifying the successful candidate for entry. Based on a range of subjects examined by a variety of means, the International Baccalaureate is intended to demonstrate achievement as well as potential. Recognising the consequences of using a test of attainment as the main predictor, authorities in a number of Commonwealth countries, including Australia, Britain and Canada, have experimented with scholastic aptitude tests. Since these do not relate closely to any particular curriculum they help to free the teacher and his pupils from the strict control of the examination syllabus. More use could be made of aptitude tests in addition to achievement tests. This approach might help to lower the undue emphasis now being placed on achievement testing. There is need to reduce the stress of examinations and the addition of aptitude tests might help alleviate of the problems. Formal assignments or project work, where a student is given a reasonable period of time to accomplish a task, might help to give students more interest and probably more insight into what they produce and may be of considerable educational benefit to the student.

Numerous techniques are being developed to cater for the detailed needs of new forms of testing. Among such techniques there may be cited the building of item banks or question libraries as resource centres from which proved material may be drawn to construct tests appropriate in form and content for the objectives decided upon. Efficient tests of this type would measure agreed objectives and yield comparable results on an achievement scale.

The objective test format has been employed in the search for unorthodox evaluation approaches which are designed to emphasize the point that many problem situations have more than one acceptable solution; in contradistinction to the one-acceptable-solution idea implicit in traditional multiple-choice tests. In one such approach, the student is required to respond to questions for which several correct options are provided. The differing options to each question reflect differing interpretations of the problem situation. In another approach, the choice is between several correct options that reflect differing levels of cognitive functioning on the hierarchy postulated in Bloom's Taxonomy of Educational Objectives - Cognitive Domain. In yet another, termed "probabilistic testing", the student is given the freedom to assign probabilities (in form of percentages to the different options according to his judgment of the extent to which each option represents a reasonable answer to the question posed.

In addition to the above, investigations involving various cognitive preference tests have been and are being conducted in attempts to find out whether low and high student achievers have some stable idiosyncratic consistencies in their information-processing strategies, which could enable their early identification for purposes of instituting instructional strategies tailored to their varying learning styles. All these new approaches, even if they are refined to the point of being very reliable and valid, cannot be regarded as panaceas. But they are welcome additions to traditional evaluation instruments in the continuing search for ways of improving the educational process through improved evaluation techniques.

While progress has undoubtedly been made in the devising of better methods of testing and measurement, certain areas still give rise to concern and call for further action. Among these there may be included the development of non-verbal and other tests designed to reduce the disadvantage of candidates being examined in languages other than their mother tongue or candidates with differing cultural backgrounds; the continuing search for more efficient predictors; the improvement of testing procedures for spoken language and practical skills; methods of examining private candidates, non-institutional candidates and mature candidates; the determination of the best forms in which to present results to those who need to use them; the role of teachers and teachers' organisations in examination development and the best means of ensuring that all teachers become familiar with basic examination techniques during their training; and the most suitable means for the initiation and control of research into new examination techniques.

Research and Evaluation in Examination Development

The search for more efficient and less constraining forms of testing benefits from being undertaken systematically. Needs must be identified and problems thrown into relief before progress becomes possible. Once pilot projects are planned, provision must be made for both ongoing monitoring (formative research) and final evaluation (summative research) if the full benefits of the experiment are to be derived and passed on to the potential consumers. Research, which may be valuable though small-scale in breadth and time, is indispensable to the effective operation of any system of testing or measurement.

Although much of the work of researchers related to examination development tends to be concentrated on the identification of improved techniques and the construction of tests and models, researchers cannot avoid the wider questions in the background. Examinations must be developed

in the context of the education system and the existing socio-economic structure. Since the background is essentially dynamic, ongoing research must continually reappraise current testing instruments. Cultural and psychological influences on candidates must not be neglected in the course of research and evaluation.

A variety of problems associated with public examinations may be brought closer to solution through research. A wide range of studies is needed to understand the adequacy of each examination, the teaching which led the pupil to it, the quality of his school experiences and a host of other variables which may be more or less relevant to the performance of the candidate. The first step is to identify the important problems and then plan and carry out a systematic programme of research. Research is needed at every stage of the examination process from the planning of the examination syllabus, the preparation and production of the questions, the marking, the grading and then the establishment of the relationship of the published grades to other aspects of education. A strong case can be made for the introduction of newer methods of assessment as supplements to traditional examinations; methods such as continuous assessment, comprehensive achievement monitoring (e.g. periodic testing), self-evaluation, item banking of common subject areas, and so on. Research can assess the effectiveness of these different methods and can also help to reveal the answers to more traditional problems such as the consistency of markers, the validity of specific examinations and the reliability of criteria against which the examinations are judged. It has been suggested that some selection problems might better be solved by lottery, owing to the abundance of qualified candidates for a limited number of positions. A lottery is an admission of the inability of the examiners to measure with sufficient precision the relevant attributes of both the candidate and the training course (or job) for which the candidate is being selected. Ideally, therefore, selection problems are better solved by gathering more information about the candidates, the predictors and the criteria.

When a choice of questions is allowed in an essay examination, research is essential to ensure that the questions are equivalent in all of their important qualities; otherwise the candidates will be taking what are effectively different examinations.

Another important problem in education is to identify the significant domains of achievement. It is eighteen years since the first volume of the Taxonomy of Educational Objectives was put forward as an approximate map of achievement. It may be that some of the areas in the taxonomy are unnecessary, while others need to be included; research can help to provide the answers.

It might appear unfortunate that many research studies are reported in precise technical terms which, although scientifically accurate and of professional quality, are not readily understood by the layman. Research reporters should recognize that their findings may be intended for more than one level of readership and that some popular summary of the main findings may be the most effective way of reporting to most readers, although the need for the precise, scientific report remains paramount.

Despite the disadvantages inherent in the situation where the same body acts as defendant and judge, on balance it seems appropriate for examining bodies to oversee the research related to their own activities because only the boards have both the data and the experience necessary for the tasks. The research team, however, might usefully include individuals

from external institutions. More basic research could probably best be undertaken by external bodies. It may be noted in this connection that not all research students following postgraduate courses can be regarded as sources of additional manpower for this purpose since their courses may be essentially a training in the use of research methodologies, precluding their engagement in the full practice of research.

In terms of cost, research and evaluation need not absorb a high proportion of overall expenditure on the administration of examinations. In the context of overall expenditure on formal education its costs will be miniscule, certainly in proportion to the effects which examinations presently exert over education systems.

In conclusion, it appears two types of research may be required. First, the "systems maintenance" type of research which is a continuing process to prepare and improve upon existing procedures and instruments throughout the examination cycle. The second may be called evaluative: the type of research aimed at discovering the effects of examinations upon the full range of social, educational, economic and political factors upon which they impinge. Efforts should be made to plan a comprehensive long range programme of research within the ambit of examining bodies on the one hand and among examining bodies in the Commonwealth on the other hand so that a comprehensive body of knowledge can be developed about the examination process.