5. ASSESSMENT OF CHILDREN'S PROGRESS AND EVALUATION OF PROGRAMMES[1]

Introduction

A few years ago it was apparently easy to examine candidates in mathematics and it seemed comparatively simple to evaluate a mathematics curriculum. At the primary school level the content of the teaching was confined to arithmetic and the debate was in terms of a conventional testing of mechanical and problem arithmetic contrasted with objective, multiple choice tests. The situation at the secondary level was equally stable. There was an interesting and valuable move away from examining mathematics in separate compartments by means of papers in algebra, geometry, arithmetic and trigonometry towards papers in mathematics as a whole.

Three developments have taken place recently or are taking place, which make the problem of examination, assessment and evaluation increasingly complex. The first is the development of new ways of examining which place much more emphasis on the assessment of course work and the opinion of the teacher.[2] The second is the development of new ways of learning and teaching which, in a number of countries, tend towards the abolition of rigid streaming by ability and the forming of classes of widely varying capability, particularly in mathematics.[3] Efficient class teaching is almost impossible in such circumstances and the good teacher is obliged to consider the formation of small groups within the larger class, and the adoption of individualised learning by means of pieces of work set for pupils to work on their own. There is support, of course, for group work and individualised teaching as conducive to active ways of learning.[4]

The third trend is the development of new content at both the primary and the secondary levels. At the primary level we are no longer content to teach the old arithmetic; mathematics has taken its place. Sometimes this mathematics takes a quite traditional form, such as simple geometry and trigonometry, or the use of graphical methods; sometimes the change is more radical in terms of sets, transformation geometry, etc. Particularly in the primary schools, the development of new mathematics and new teaching and learning procedures are often combined together in such a way that their effects cannot be separated. In the secondary schools, so far, the change has been, in the main but not entirely, towards new content.

The fourth trend, especially in developing countries, is the phenomenal rise in the numbers of candidates. Mathematics in one form or another is a subject of the selection examinations into the secondary school and one of the subjects offered by most candidates in the school certificate examination. This trend is responsible for an increasing automation in the marking of students' examination scripts.

---

[1] Adapted from the lead paper by Professor J. Wrigley.

[2] Secondary School Examinations Council: The Certificate of Secondary Education - some suggestions for teachers and examiners, London, H.M.S.O. 1963.

[3] cf. Comprehensive Schools.

[4] E.A. Peel: The Psychological Basis of Education, London, Oliver and Boyd, 1964.

All these developments lead to considerable problems in examination and evaluation. Although the examination of candidates and the evaluation of curricula are often considered as separate problems, a little consideration readily shows that they are inextricably interwoven. An examination first differentiates between the candidates and then fixes a general standard of achievement. If we are examining pupils who have been taught to a new and radically different syllabus, it is comparatively easy to grade the pupils in an order of merit (the first function of an examination), but it is much more difficult to make judgments about standards. Questions of comparability with the old standards arise; sometimes they can hardly be immediately answered. As soon as we make statements about standards we are also evaluating the curriculum as well as the teaching and the pupils. So our present task with regard to examination, assessment and evaluation is at once both more interesting and more difficult than it used to be. And in the immediate future the difficulties are likely to increase rather than be resolved.

Examinations

External examinations have played an important part in the development and progress of many countries. The trouble with most examinations is that we ask of them too much. We expect them to differentiate between candidates, to provide evidence of standards, to act as guides to good teaching, and to provide incentives for both pupils and teachers. We expect them to predict future performance as well as to certify that a candidate has completed satisfactorily a course of recognised study. Add to all this the fact that we examine candidates in their tens of thousands and it is not surprising that we run into difficulties. Ideally, we should separate these various functions, asking ourselves exactly what we aim to do, and design instruments to carry out our aims. But we are rarely able to do this.

A most interesting development in the realm of examining in recent years is the examination for the Certificate of Secondary Education (C.S.E.)[1] in England and Wales. It was set up to provide an examination suitable to the needs of pupils in the secondary schools, who were not in the top ability groups. It was designed roughly for pupils between the 80th and 40th percentile of the ability grouping. Thus it was of a lower standard than the General Certificate of Education (G.C.E.) Ordinary Level and intended to suit average children and above. It was to take the place of a number of external examinations which had been taken by these pupils and which had become rather remote from the needs of the pupils in school. To ensure that the new examination was relevant, teachers were put in control of all the important committees and fourteen C.S.E. Boards were set up throughout the country. The philosophy of C.S.E. is that teachers and examiners should coincide, should often be the same people, that a teacher knows best the capability and calibre of his own pupils but that he needs the help and guidance of external examiners in the final determination of national standards. For many subjects, and for pupils of average ability, it was clear that conventional methods of examination were no longer adequate, so the C.S.E. was designed so that it could be taken in three different ways or modes. Mode I is a conventional external examination, Mode II is an examination set externally on a syllabus designed by the pupils' own school and teacher, Mode III is an internal examination externally moderated. The latter is a means by which an enterprising teacher can examine his own work

[1] Secondary School Examinations Council: The Certificate of Secondary Education - some suggestions for teachers and examiners, London, H.M.S.O., 1963.

and his own syllabus in the way he thinks best within his own class, though of course he is subject to the checks of an external moderator. The C. S. E. Board has the final responsibility of making sure that work under Mode III is comparable in scope and in standard with work under either Mode 1 or Mode II.

The development of Mode III examinations with the associated trend to continuous assessment and the evaluation of course and project work has been uneven throughout the country and has depended upon the policy of the individual examining boards. Certain subjects of the curriculum such as rural studies, home economics, music seemed to cry out for the imaginative use of Mode III procedures, but a similar development has taken place in the more traditional and academic subjects, and mathematics has been one of these. In particular, those teachers who have been quick to see the value of new content in mathematics and/or the open-ended way of teaching have wanted to experiment with Mode III. The intention of many of these people has been to try to reproduce, albeit at a lower level, the kind of activity of a professional mathematician. This demands project work, essay-type questions, open-ended situations, and a whole new style of teaching, learning and examining. The number of these vanguard teachers is not necessarily large but it is growing. The problems raised for the moderator in such a situation are considerable. How is he to make sure that standards have been maintained, that the subject taught is still recognisably mathematics? Given tact and understanding, a reasonable assessment and evaluation can take place. The teacher must be trusted to assess his own pupils to a large extent especially as regards an order of merit within the class. But the moderator might expect, even with the most esoteric subject matter and the most free and easy learning situations, that on some key aspects of mathematics traditional questions could be asked. It is evident, though, that this kind of compromise does not completely solve the problem of comparability - the trustworthiness of the teacher and the wisdom of the moderator are essential elements in the situation.

The new ideas of teachers and examiners coming together in one person, of ratification of progress rather than external examination, of continuous assessment rather than a once-and-for-all external examination, of the assessment of course and project work are present to a greater or lesser extent in all the three Modes of examination and in all the C. S. E. Boards. A system of internal examinations, continuous assessment, project work, open-book type examinations has not yet been extensively tried out in highly competitive situations. With most of these ways of examining, the interaction between teacher and pupils becomes rather obvious. In examining a dissertation or a long essay the influence of the tutor on the work is considerable and it is sometimes difficult to know how to allow for this to give a 'fair' assessment. The fact that a similar situation exists in an external conventional examination (anyone who has examined large numbers of mathematics scripts at G. C. E. 'A' and 'O' levels will know that he is examining schools and teachers as much as candidates) does not mean that we have a solution to the problem.

The kind of development in examining that has been described is comparatively sophisticated. It takes the risk of bringing in the teacher as an examiner, it admits openly the possibility of bias; it is difficult to accomplish in a highly competitive situation. Such methods will flourish best in a system which provides expanding opportunities for education where the effects of failure can be retrieved at a later date. It is an interesting question as to whether a similar movement is suitable for developing

countries. Since these new ways of examining (at their best) lead to more relevant and less artificial teaching and learning perhaps they are all the more essential. But they demand well qualified teachers who know exactly where they are going and what they are doing. The external London G.C.E. (and even degree examinations) set a standard which is widely recognised, but sometimes at the expense of some unreality. Perhaps a compromise is necessary. Part of the examination in mathematics could be a conventional two or three hour written paper; this would take care of the essentials and provide an external yardstick. Part could be an assessment of course work, and open-ended examination - this would encourage real mathematics in the schools.

A development of examinations is the objective-type as distinguished from the conventional type which demands written statement, step by step, leading to the answer. An examination, either the objective-type or the conventional type, is devised to assess the attainment and skill of pupils in a particular subject. All the questions refer to a syllabus defined for the class concerned. An important difference between the objective and the conventional written examination is that the former usually consists of a large number of questions which cover the syllabus more extensively and thoroughly. In the conventional examination, the questions are few and the coverage is not as systematic and extensive. There is therefore a greater chance that the questions included will suit some pupils very well and others less so, in spite of the practice in some papers of allowing candidates to select a few questions from the paper. The issue does not arise in an objective examination where candidates are expected to answer all questions, which cover the whole syllabus.

The objective-type examination has been viewed with suspicion. Some doubt if one-word answers can ever be an adequate form of assessment and think that even if it were efficient the backwash effect on the teaching would be bad. The second objection is true of many external examinations, but the first is not true if the examination is well constructed. Good objective-type examinations can examine high level modes of thinking efficiently. The pressure of examining large numbers of pupils will force us in one of two directions either towards more teacher and school-based assessments or towards more multiple choice tests which can be scored by machine. Many examining boards now use some objective-type in school examinations, such as the selection examinations for admission into secondary schools and the school certificate examinations comparable to the G.C.E. Ordinary level.

The most usual form of the objective-type examination is the multi-choice in which a candidate is set the task of selecting a current answer from a list of alternative answers offered. A later development is the multi-facet form in which a mathematical situation with many facets is presented and the candidate is required to make a decision on each of the facets. In exploiting this multi-facet idea statements were phrased so that some were true and some were false and the candidates were required to decide the truth or falseness of each statement made about each situation. Here are two examples of the multi-facet form of the objective-type examination:

(1) If $a = \frac{1}{2}$, $b = \frac{2}{3}$ and $c = \frac{3}{4}$, then*

(A) $abc = a^2$

(B) $a + b + c = \frac{6}{9}$

(C) $b - c = \frac{a}{6}$

(D) $2(c-a) = 3(b-a)$

(E) None of the above is true

	True	False
A	x	
B		x
C		x
D	x	
E		x

(2) S_1 and S_2 are concentric circles with radii 3 cm and 5 cm**

(A) The area of S_1 is $\frac{3}{5}$ of the area of S_2

(B) The circumference of S_2 is $\frac{3}{5}$ of the circumference of S_1

(C) The circumference of S_1 twice the diameter of S_2

(D) All tangents to S_2 are chords of S_1

(E) Any chord of S_2 which is tangent to S_1 is 8 cm. long

	True	False
A		x
B		x
C	x	
D		x
E	x	

One of the criticisms of the method was that there was a guessing element involved in True-False answers. These criticisms were not entirely met by scoring the results of the test by means of a formula involving Right-Wrong answers. However, the True/False element in the multi-facet situation is not essential - multi-facet questions could just as easily be phrased in such a way as to require the answers to be found and stated in a conventional way. For example***

(1) You are given the following information:

$a = \frac{1}{6}$, $b = 0.60$; $c = \frac{2}{9}$; $d = 0.56$;

$e = \frac{1}{4}$; $f = 0.64$

(A) Place a, b, c, d, e, f, in order of size, the greatest first.

(B) Find the value of (c-a)-(e-c)

Answers

*Secondary School Examinations Council: The Certificate of Secondary Education: Experimental Examinations - Mathematics (Examinations Bulletin No. 2), London, H.M.S.O., 1964, p. 20.

**Ibid.

***Bulletin No. 7.

(C) Is the average value of b,d, and f equal to, greater than or less than the average value of d and f?

(D) How great is the difference between $\frac{a}{c}$ and $\frac{d}{f}$?

(E) What is the value of the product abd?

Answers

The Psychometric Movement

The word 'test' has been avoided in describing the objective-type examination because, in addition to its usual meaning of an examination, it has a technical meaning[1] of a set of multi-choice and multi-facet objective-type questions constructed by persons technically trained in mental testing, statistics and psychometric techniques. The items of the test are thoroughly tried out and standardised before the test is administered to candidates. A test may be either psychological or educational. A psychological test assesses such abilities as general intelligence, verbal reasoning, spatial judgment, mechanical aptitude and so on. An educational test assesses achievement in subjects or technical skills. An educational test looks like what any teacher might prepare as an objective-type examination. In fact, a standardised test requires time and technique for construction. Measures of item difficulty and item discrimination enable the test constructor to construct tests which are technically efficient in that they discriminate between candidates and they are reasonably reliable and consistent. Providing the test constructor also considers carefully his aims and objectives, prepares a blue-print which makes sure that his aims are realised through an adequate sampling of the course of instruction, a thoroughly efficient test can be constructed. It is however very important to keep in the forefront of one's minds the aim of the test. If it is required to differentiate between candidates over a large range of ability then each item should discriminate. In a short test we cannot afford the luxury of items which all candidates either answer correctly or incorrectly. Hence the concern of the test constructors for item statistics. But not all tests, and certainly not all examinations, are of this kind. A teacher may need to use a test to see whether what he has taught has been understood and assimilated. In particular, there will be certain key operations which must be performed automatically and accurately. In the stress on teaching for understanding the need for automatic and accurate response is sometimes forgotten. In testing for such qualities we want a completely correct response from all our pupils. We are thus quite content to have items which do not discriminate and do not have any incorrect answers. In this situation, which ought to be common in teaching, item statistics are almost irrelevant. Similarly in curriculum evaluation we may be interested to know whether or not the children have understood key concepts. Here we may need items which measure understanding of these concepts. The items assembled for this purpose may be of the kind in which the response is either correct or incorrect and a whole test could be constructed so that something like a 90% correct response is required if the teaching and learning situation can be considered

[1]Vernon, P.E.: The Certificate of Secondary Education - an introduction to objective-type examinations (Examinations Bulletin No.4), London H.M:S.O., 1964, p.2.

satisfactory. Similar considerations apply in programmed learning. The moral of all this is that the whole apparatus and theory of mental testing must be looked at carefully and the aims of the test or the examination must never be forgotten.

The Test Development and Research Office of the West African Examinations Council, Lagos, Nigeria, constructs tests in mathematics and other subjects for the selection examinations and the West African School Certificate and the G.C.E. examinations.

Curriculum Evaluation

The pattern of curriculum development in both the United States and the United Kingdom over the past decade has been one of acts of faith and trial and error. This has been particularly so in mathematics and science. Men and women of ability and vision have become increasingly dissatisfied with the content of traditional mathematics both as regards its intrinsic mathematical value in this modern age and for its impact in terms of relevance, interest and difficulty for the pupils in the schools. They have had an unshakeable belief that new content should be introduced, and they have gone ahead and experimented with it. Sometimes the most interesting new ideas and new content have been suggested by professional mathematicians with little experience in the classroom. In Great Britain, the unreality which might stem from a lack of practical experience of children has been avoided by organising elaborate and extensive trials of the new materials and the new methods in school. Many teachers and many schools have been brought in to the developments, and arrangements were made by feed-back procedures to ensure that the lessons learned were used to modify the new materials. There has been used a system of trial and error that is in itself a kind of curriculum evaluation which should not be despised or undervalued. There is certainly need to go a good deal further. Curriculum development projects cost money and it is essential to know whether the money has been well spent and how to spend money more efficiently on the next project. It is important that the effects of the new curricula on the lives and experiences of the children are healthy.

In a new development in curriculum, the general and the particular aims and objectives of the development should be set out. The specific objectives which follow from the general aims should be interpreted in terms of behaviour. Once the behavioural changes are known, test situations can be devised to measure the change. Thus the success of the curriculum in terms of the developers' own aims and objectives can be measured. It is then a separate issue, best done independently, to evaluate the intrinsic value of the original intention of the project as a whole. It is becoming usual to attach to any curriculum development team a psychometrician whose role is to ask pertinent questions of his colleagues, to get them to clarify their aims and objectives, to organise tests of pupils' understanding and of changes in behaviour, and to make sure that the feed-back of information derived from trials is efficient, adequate, and acted upon. Curriculum development is becoming a feature of educational development in many countries. It is necessary to evaluate the new curricula and those engaged in these projects would find useful the experience of similar project teams, such as the Nuffield Ordinary Level G.C.E. Science Development and the Nuffield Primary Mathematics Project.

Conclusion

Examinations, tests, assessments, and the evaluation of curricula are all inter-related. In assessing and examining children we are to some extent evaluating our curricula, new and old, and we are also measuring our success as teachers. Examinations, both external and internal, still have an important part to play in the maintenance of standards, in the guidance of teachers, perhaps in acting as incentives for pupils, and for qualifying and selection purposes. The best examinations in the future will be a compromise between the 'teach and test' procedures which all good teachers need to employ, and the external checks which are still necessary. The new examinations must not be allowed to inhibit the most enterprising teaching and they must allow intrinsic good teaching of all that is best in mathematics to take place. The best psychometric procedures should be adopted to make the examination as reliable and valid as possible. In curriculum evaluation, we should strive to make our evaluations as scientific as possible but should recognise also the merit of rough and ready procedures. What we need are inspired examiners and imaginative evaluators. The most hopeful long term solution to the problem is to make the teachers and the testers coincide in the same people, and similarly to combine the developers and the evaluators if not into the same people at least into the same team of people.

Suggestions for Action

1. Whatever the programme, it is important for the teacher to assess the progress of the pupils, test their knowledge and skills from time to time and as far as possible ascertain their potential and attitudes. Therefore the teacher must keep himself up-to-date in the methods of examinations, tests and assessments of pupils and evaluation of programmes.

2. In order that the assessment of pupils may be meaningful, the teacher should understand the curriculum and its objectives and preferably participate in the construction and review. He should therefore demonstrate concern and interest by making the assessment and evaluation and feeding back to the curriculum development team.

3. Teachers' professional associations can influence examination practices and curriculum changes. Therefore a teacher should belong to his professional association.

4. Teachers generally and primary school teachers in particular must try to ensure that the child is ready to go on to the next step or to a new experience.

5. It is important that an understanding of mathematical language be properly assessed, especially with young children.

6. Continual evaluation by the teacher is of the utmost importance and records should be kept. Not only will this help the teacher to assess the pupils but also to assess his own procedures.

7. Where a qualifying or selective test has to be set externally it is recommended that teachers should be involved as much as is possible in devising the content and method of examination.

8. In countries where a selective or an achievement test is necessary, such a test should include multiple-choice, multi-facet, and traditional types of items. Where possible, course work and practical tests should also be included. Where teachers are able, it is desirable that teacher opinion be given high rating in the assessment of students.

9. Overall aims, as well as aims within stages of a programme, should be clearly stated and the programme or project can then be evaluated in terms of those objectives. A sharing of opinions about and comparison of results from similar programmes in different countries is to be encouraged. For this purpose, a teacher or his school should subscribe to one or two mathematical journals from other countries.

10. Teachers should avail themselves of in-service courses in examinations, tests, assessments and evaluation techniques.

Selected Reading List

1. Board of Education: Curriculum and Examinations in Secondary Schools, London, H.M.S.O., 1946.

2. Bruce, George: Secondary School Examinations Facts and Commentary, London, Pergamon, 1969.

3. Educational Testing Service: Tests and Measurements Kit, Princeton, New Jersey, Educational Testing Service, 1970.

4. Fleming, C.M.: Cumulative Records, London, University of London Press, 1946.

5. Joan, Dean: Recording Children's Progress, Basingstoke, Macmillan Education, 1972.

6. Mathematical Association: Transfer from Primary to Secondary Schools, London, G. Bell and Sons Ltd., 1964.

7. Pidgeon, D. A.: Evaluation of Achievement, Basingstoke, Macmillan Education, 1972.

8. Secondary School Examinations Council: The Certificate of Secondary Education - An Introduction to some Techniques of Examining, London, H.M.S.O. 1964.

9. Secondary School Examinations Council: The Certificate of Secondary Education Mathematics, (Bulletin No.2), London, H.M.S.O., 1964.

10. Secondary School Examinations Council: The Certificate of Secondary Education, Mathematics 2, (Bulletin No. 7), London, H.M.S.O., 1965.

11. Secondary School Examinations Council: The Certificate of Secondary Education - some suggestions for teachers and examiners, London, H.M.S.O., 1963.

12. The Schools Council: The Certificate of Secondary Education - School-based Examinations: Examining, Assessing and Moderating by Teachers (Examinations Bulletin No.5), London, H.M.S.O., 1965.

13. Vernon, Philip E.: The Certificate of Secondary Education - An Introduction to Objective-Type Examinations (Examinations Bulletin No.4), London, H.M.S.O., 1964.