CHAPTER VI

Assessment of Children's Progress, and Evaluation of Programmes: Purpose and Method

Lead paper by Professor J. Wrigley, B.Sc., M.Ed., Ph.D., F.B.Ps.S., University of Reading, England.

Introduction

1. A few years ago it was apparently easy to examine candidates in mathematics and it seemed comparatively simple to evaluate the mathematics curriculum. Twenty years ago at the primary school level the content of the teaching was confined to arithmetic and the debate was in terms of conventional testing of mechanical and problem arithmetic contrasted with objective, multiple choice tests. The latter could be demonstrated to be more consistent and reliable than the former but both kinds of examination in mathematics appeared to be more reliable than similar procedures in other subjects, particularly English. 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. Even this highly desirable change took many years to come to full fruition in conservative England. At the eighteen year old level even less change took place until very recently.

2. The situation today in England is somewhat different and will become even more complex in the years ahead. Three developments have taken place recently, all in my judgement desirable, but all calculated to make the problem of examination, assessment, and evaluation more 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. In England these developments have taken place around the new examination taken at the age of sixteen years, the Certificate of Secondary Education (C.S.E.). The change has come with regard to all subjects, but mathematics has been at the forefront in the various experiments which have taken place. The second is the development of new ways of learning and teaching which have come into existence in the primary schools and which, before long, will begin to spread upwards to the secondary schools. The trend is towards the abolition of rigid streaming by ability and the forming of classes of widely varying capability, particularly in mathematics. Efficient class teaching is almost impossible in such circumstances and the good teacher is forced to consider the formation of small groups within the larger class, and even the adoption of individualised learning by means of pieces of

work set for pupils to work on their own. Although I say that the teacher is forced to adopt these methods this does not mean that there are not good pedagogical and psychological reasons for adopting more active ways of learning. There are, and they are well demonstrated in Curriculum Bulletin No.1, 'Mathematics in Primary Schools' (1965 H.M.S.O.).

3. The third trend is the development of new content at both the primary and 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 our 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. We have now, in England, well established new courses in mathematics, the Schools Mathematics Project, the Midlands Mathematics Experiment, the Mathematics in Education and Industry etc. A real breakthrough in examining came when the S.M.P. persuaded the G.C.E. examining boards to examine its work in its own terms. There was no questions of examining the pupils who had followed a new course by the examinations set for traditional courses. It seems obvious enough now but at the time it was a considerable step forward. More and more special courses, in many other subjects as well as in mathematics, are being developed and are being examined in their own right by special papers.

4. All these developments lead to considerable problems in examination and evaluation. Although the examination of candidates and the evaluation of curriculum 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 judgements 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

5. For a long time in the United Kingdom examinations of a conventional kind have played an important role in ensuring the maintenance and enhancement of standards and in providing an external measuring rod which could be seen to be fair to all candidates from whatever strata of society they came. The examinations at 11+, 16, and 18 have all played an important part in the democratisation of society and in making it possible for all children in our country to have opportunity for advancement. The fact that we now know that the opportunities have never been equal does

not alter the fact that external examinations have played an important part in encouraging social mobility. 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.

6. The most interesting development in the realm of examining in recent years in England and Wales is the advent of the C.S.E. It was set up a few years ago to provide an examination suitable to the needs of pupils in our 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 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 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 I or Mode II.

7. 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 *avant-garde* 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? My own view is that 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 aspects of mathematics traditional questions could be asked about certain key aspects of mathematics. 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.

8. I have been discussing rather extreme examples. 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. Similar procedures have been adopted in College of Education examinations for a long time. It remains to be seen whether a similar movement will take place in the universities and at the Advanced level of G.C.E. 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 pupil 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.

9. The kind of development in examining that I have been trying to describe is comparatively sophisticated. It takes the risk of bringing in the teacher as an examiner, it admits openly the possibility of bias. It would perhaps have been an impossible development 50 years ago in the United Kingdom, it is still 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 under-developed and emerging 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 for the whole Commonwealth, 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 our schools.

The Psychometric movement

10. Examinations and tests are both forms of assessment. Yet in England there has been until recently a sharp division between the two. Objective multiple choice tests have been viewed with suspicion by many of the people who operate our examination system. There are two main reasons for this suspicion – first that objective testing by means of one word answers cannot ever be an adequate form of assessment and second that even if the multiple choice activity were efficient as a test procedure the backwash effect on the teaching would be almost wholly bad. Memories of the coaching for the eleven plus examinations die hard in England. The first reason for suspicion is in my view not well founded. Good objective tests can examine even high level modes of thinking efficiently. The second reason is much more soundly based – until the competitive element in our examinations is reduced the backwash effect of any examination will be considerable – if we are to make it beneficial we may have to continue with the kind of interesting innovations I have described in the last section. 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. The examining boards in England are beginning to experiment with objective tests both in the C.S.E. and G.C.E. sectors.

11. Although my personal view is inclined towards the development of more school based examinations yet there are some interesting developments in the realm of multiple choice examinations. In mathematics one of the most interesting has been with regard to 'multi-facet' tests as opposed to multiple-choice ones. The idea is fully described in Examinations Bulletins No.2 and No.7 – The Certificate of Secondary Education: Experimental Examinations – Mathematics – *Mathematics (Bulletin No. 2)* H.M.S.O. 1964. *Mathematics 2 (Bulletin No. 7)* H.M.S.O. 1965. The normal multiple choice situation sets a candidate the task of selecting a correct answer or answers from among the alternatives offered. Such tests derive only one question from a given situation and therefore call for only one independent decision. It is however frequently possible to consider the same mathematical situation from a variety of facets and consequently to require from it a variety of decisions. 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.

Two examples will make the method clear:-

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

(A) $abc = a^2$

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

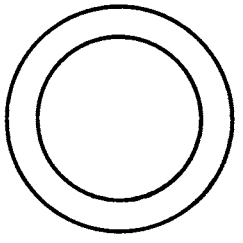
(C) $b - c = \frac{a}{6}$

(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) C and S are concentric circles with radii 3 cm and 5 cm



(A) The area of C is $\frac{3}{5}$ of the area of S

(B) The circumference of S is $\frac{3}{5}$ of the circumference of C

(C) The circumference of C < twice the diameter of S

(D) All tangents to S are chords of C

(E) Any chord of S which is tangent to C is 8 cm long.

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

12. The decision making situation combined with the obvious mathematical quality of the test described in the Examination Bulletins led to the method being adopted quite widely in England in the C.S.E. examinations. 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 (from Examinations Bulletin No.7):-

(1) *Situation* You are given the following information:

$a = \frac{1}{6}$; $b = 0{\cdot}60$; $c = \frac{2}{9}$; $d = 0{\cdot}56$; $e = \frac{1}{4}$; $f = 0{\cdot}64$

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

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

(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?

Answer

13. Psychometric techniques have been developed for measuring the performance of items in tests. Good tests are tried out before actual use and item statistics calculated. 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 blueprint

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 mind 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 things in any mathematics syllabus which must be understood and 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 irrelevent. 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 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.

Curriculum Evaluation

14. Quite the most difficult task of those of us interested in assessment and evaluation is to evolve ways of evaluating new curricula. Whatever the problems involved in testing and examining children in traditional situations there is a vast store of sound theory and good practice which has been built up over the years. But the question of how to evaluate a new curriculum has hardly been tackled as yet. Indeed some of our most inspired curriculum developers seem unable to accept the need for evaluation except in a very rough and ready fashion. 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 on the pupils in our 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 however we have avoided the unreality which might stem from a lack of practical experience of children by organising elaborate and extensive trials of the new materials and the new methods in school. We have brought in to the developments many teachers and many schools and made sure by feed-back

procedures that the lessons learned were used to modify the new materials. We have used a system of trial and error that is in itself a kind of curriculum evaluation which should not be despised or under-valued. Nevertheless we certainly need to go a good deal further. Curriculum development projects cost money and we need to know whether the money has been well spent and how to spend our money more efficiently on the next project. More importantly though, children's lives and experiences are affected by what they learn; how can we be sure that the new curricula are worthwhile? So far we have hired the best men available and given them trial and error apparatus to make sure that they can check their hunches and their practices. Can we do more?

15. The conventional answer to this question is as follows. The aims and objectives of a new development in curriculum should be set out, first in general and then in particular. The specific objectives which follow from the general aims should be able to be interpreted in terms of behaviour. "Down with non-behavioural objectives" is a slogan which has been seen on both sides of the Atlantic. 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. Such a programme for curriculum evaluation is logical – it might seem quite appropriate in a subject such as mathematics. It is however highly unpopular with many teachers and with many creative curriculum constructors. It has rarely been carried out in its entirety. The reason for the opposition is two-fold. First, the most creative minds simply do not work in this logical, analytical fashion. If other minds are set to work alongside the creative developers tensions immediately form. Second, the logical process of analysis seem to exclude a number of important intangible things which people feel to be important – the flair of the teacher, the hunch of the developer, the interest of the child – good teaching just does not work in the logical way suggested above.

16. In Britain the best we have been able to do so far is to develop a distinction between on-going and final evaluation. We are beginning to attach to any curriculum development team an individual 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. Ideally such a man would be skilled in measurement techniques and sympathetic to the aims of the project. Ideally the tension which might develop between the evaluator and the developer would be healthy – this is most likely to occur when the individuals work closely together.

17. We have made less progress with large-scale independent evaluation. For the Nuffield Ordinary level G.C.E. Science development a research project has been set up to produce instruments to measure the knowledge and ability of pupils in science at the appropriate level. Such instruments, chiefly in the form of objective tests, have been designed to measure performance in science of pupils who have been subjected to both traditional and new science schemes. It is in this connection that

the different kinds of test construction mentioned previously can be employed. The tests will not be used to simply differentiate children's ability, nor will they be used to compare total score on the test between groups. What will be important is the pattern of scores obtained for differing sections of the test – both in terms of content and in terms of understanding of concepts. The instruments, when finally constructed, could be used as part of a national evaluation of science teaching in the country.

18. Two other developments are worth mentioning. For the Nuffield Primary Mathematics Project the assistance of the Institut des Sciences de l'Education at Geneva using Piaget-type test situations has been enlisted. A series of "Check Ups" have been designed specially for the project. The intention is to test the pupil individually in a reasonably well-structured and standardised situation to see if he has really understood a particular concept. The check-ups are not tests and there is no question of 'right' or 'wrong' marking. They are given to the children as and when the teacher feels the need to know how far their thinking has progressed. The methods follow the usual Piaget situation and contribute an interesting example of curriculum evaluation. Nevertheless they cannot be regarded as an independent final evaluation designed to see if the project is carrying out its aims and objectives. Rather they help the on-going evaluation and provide evidence of satisfactory progress on the part of the children.

19. Some people in Britain are worried in case the 'Check-ups', like examinations and tests before them, should be abused. One hears stories of the check-up situations being mistaken for the teaching and learning situation. The teacher could be content to teach simply for success in the check-up situation. But anything can be abused – the pressure on check-ups should surely not be particularly severe in a primary school situation in which the competitive element in mathematics is less than it has ever been before. And the procedure remains a most interesting and promising development in evaluating the success of teaching and learning in an intrinsic and fundamental way via the mastery of concepts.

Conclusion

20. The underlying theme of this paper is that 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. Traditional and modern mathematics exist side by side – indeed all sensible people are already synthesising the two. For our traditional procedures we had almost forgotten the need to evaluate their intrinsic worth, rather taking it for granted. We had simply examined children, relying on accumulated wisdom to set our standards. In the new situation the two tasks loom large on our horizon. Our problem is complicated by a double change – first in terms of new content and second in the development of different modes of learning. 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.

21. Similarly, in terms of curriculum evaluation we are only at the beginning of inventing new procedures. Although we should strive to make our evaluations as scientific as possible we must also recognise the merit of rough and ready procedures. It may be that we could employ a whole series of measures to determine the success of our work, some of which at first sight have only a tangential relationship to our work. For example, measures of truancy, lateness, disobedience, proportion of children staying at school, increasing (or decreasing) numbers studying mathematics, threatening statements in the classroom, delinquency, numbers of books read etc. etc. All these could be quantified; taken together, with additions, the sum total could add up to a kind of evaluation.

22. The tension between examiners, testers, and evaluators on the one hand with inspired teachers and dedicated curriculum developers on the other is perhaps inevitable. What we need are inspired examiners and imaginative evaluators and such people are in short supply. The most hopeful long term solution to our problem is to make the teachers and the testers coincide in the same people, and to similarly combine the developers and the evaluators if not into the same people at least into the same team of people.

23. In mathematics, we have now had a good run for the development of new content and new ways of learning. Much remains to be done in terms of extending the ideas and the methods to more pupils and more teachers. And much remains in rationalising, analysing, and sifting the best in the new. In doing just this we examine, assess and evaluate. We must take care not to become arid in the process.

REPORT OF WORKING GROUP B.1

Chairman: Mr. Victor Matthews (Barbados)

Introduction

24. Teachers of mathematics are challenged to have higher aims. Refinement of aims affects the methods both of teaching and of assessment.

25. It is important that every "unit of teaching" should have an aim. It is appreciated that in a programme of experimentation the aims can be revised as the children discover unexpected patterns, relationships etc. However, children may be involved in a great deal of activity, yet it does not necessarily follow that they are learning any mathematics.

26. Whatever the programme, it is indeed important for the teacher to assess the progress of the pupils. The children's knowledge and skills must be tested from time to time. Their individual potential and attitudes should as far as possible be ascertained.

The Classroom

27. The teacher's checks or tests should help him to avoid wasting his own or the pupils' time. He must try to ensure that the child is ready to go on to the next step or to a new experience.

28. Frequent informal assessment on the part of the teacher may be one way of cutting down on too much formal assessment. However, from time to time the evaluation must be both searching and objective.

29. Opportunity should be provided for each pupil to measure up to something – the objective is for each pupil to achieve as much as he is able. One should take into account all aspects of the situation e.g. the total performance of the pupil in the school's overall curriculum.

30. A good teacher-pupil relationship is important in assessment so that the interest and progress of the individual pupil may not be impeded.

31. There should be a continuing evaluation of all pupils and programmes.

Tests and Checks

32. We should rely on the expertise and knowledge of the Piaget School as their ideas are relevant to the problem of trying to evaluate the level of a pupil's understanding. The checks used in the Nuffield Project are useful in this context. These will however have to be adapted for use in other countries. In particular they must be based on a good level of understanding of the language used between teacher and pupil. The introduction of Piaget-type testing should be gradual. Confidence in using these methods will vary according to the individual. In-service training of teachers should include testing procedures. The development of the use of such checks in a school could be helped by a teaming of teachers.

33. We have advocated a continuous evaluation of pupils. This evaluation will include verbal and written tests some of which may be competitive. However, a pupil is too often evaluated on a poor final paper his early enthusiasm or intermediate better performance gaining him no credit. Comments on course work should be recorded and not just the marks gained in tests. Work done on a unit or a project should be noted.

34. Easily tested topics or repetition of experiences which are easy to create lead to boredom. This could cause an apparent falling off in the child's performance.

35. Parents may be worried about the value of new approaches in the teaching of young children. The school, through such organisations as Parent-Teacher associations, should help parents to understand modern methods and the attitudes of new programmes. Experiences at home and school can be complementary.

Selection

36. Some form of selection of pupils on the basis of examination exists in all countries. Some countries have been able to postpone this type of selection until the

end of the Secondary School Course. In most developing countries, however there are limited facilities for secondary education. Appreciating the need to encourage social mobility these countries have to select pupils for secondary level, usually at 11+.

37. The type of examination used for selection varies according to the country's needs. By and large, however, they are externally set although there are examples in some countries of teachers' control in the design and content of these examinations. The prevalent tendency is to use standardised tests capable of being machine-scored.

38. There is no doubt that such examinations have considerable influence upon the teaching in the schools. It is most desirable that teachers and external examiners should work together in devising the content and method of such examinations.

39. It is desirable that teachers' opinion should be considered and that a verbal test suited to the particular country be used along with the mathematical test.

Purpose and Method

40. The Group considered the various purposes and methods of testing. Time was spent ranking the suitability of various types of tests for serving different objectives.

Specially considered were the following purposes:

Diagnosis, Selection and Prediction,
Motivation, Qualification and Speed

and the following types of test or assessment:

Multiple-Choice, Multi-Facet, Oral Formal,
Traditional, Practical, Open ended, Course Work,
Open Book, Teacher Opinion.

Examples of examinations which require tests for qualification are those for which a certificate is awarded (e.g. G.C.E.). The Group considered that types of testing preferred for selection were not the same as those preferred for 'motivating'.

41. There was consensus that:

Teachers should make use of a wide variety of types of tests or test items in assessing their students and that where teachers are able it is desirable that teacher opinion be given high rating in the external assessment of students for achievement or selection.

42. It was reiterated time and again that it was important to have *in-service courses* for teachers and that such courses should pay regard to problems of assessment. Enthusiasm alone is not enough.

Secondary Level

43. The types of tests used in the new programmes at Secondary level were considered and the trend to involve a multiple-choice test was noted.

44. It was felt that the multi-facet type of question would help to cover a good deal of ground and thinking but that ideally examinations should not be of one form. The majority felt that there must be a time limit on all written tests. The C.S.E. type

of examination in Britain was favourably considered and it was felt that teacher-involvement should be encouraged in examinations with similar objectives in other countries.

45. A delegate from one country described an examination system up to the Matriculation level in one State which depended almost entirely on internal assessment. This was regarded as highly desirable but it was recognised that for many countries it was as yet an impossible ideal. Throughout this country scholarships to enable pupils to complete their secondary education were awarded on the results of an examination involving a paper on Quantitative Thinking with consideration given to school assessment.

46. The involvement of Course Work in assessment was considered most appropriate in a stable situation.

Evaluation of New Programmes and Curricula

47. We are challenged to ensure that new curricula are satisfactory. The expense of development projects and more importantly welfare of the children involved must be considered. The group concerned with the assessment and evaluation of New Programmes submitted that 'New' and 'Old' Programmes cannot be easily compared. The objectives are different. Each programme can however be assessed in terms of its own specific objectives.

48. If teachers in the classrooms are involved in developing and planning the new programme, other teachers will be more confident in adopting it.

49. Teacher opinion is of prime importance. The need to evaluate continually and to modify where necessary cannot be too highly stressed. The opinions of teachers will need to be supplemented by tests and examinations from time to time. These tests must be carefully devised and it is necessary that teachers and evaluators should clearly understand how to proceed.

50. In certain projects central schools may be carefully chosen to take some of the tests. There are examples of this kind of evaluation on both sides of the Atlantic. Standards can be checked by a test including items based on mathematical content common to both programmes in conjunction with some kind of independent parameter such as an intelligence test or a paper on quantitive thinking. Checks made in project schools and control schools have also attempted to compare the level of pupils' interest and the attitudes of teachers.

Conclusion

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

52. Continuous 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.

53. The value of trying to ascertain the level of understanding of pupils cannot be underestimated. Piaget-type checks should be given to pupils when the teacher needs

to know how far their thinking has progressed. Other forms of evaluation, formal and informal, must still be applied. Traditional type tests with marking schemes awarding marks for method and accuracy are still considered appropriate for testing achievement although it is doubtful how they affect motivation. We recommend that teachers use a variety of tests in assessing their students.

54. 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. A cold assessment of new programmes must be done at some time.

55. 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.

56. Where a qualifying or selective test has to be set externally we have recommended that teachers should be involved as much as is possible in devising the content and method of examination. It is encouraging to note that CREDO has offered to assist countries in problems of examination for selection.

57. Finally it is desirable that there be a further sharing of expertise and literature about assessment and evaluation between countries of the Commonwealth.

PLENARY DISCUSSION OF REPORT OF WORKING GROUP B.1

Survey and Comments

58. Discussion centred around the statement in Paragraph 25 "It is important that every unit of teaching should have an aim". Some delegates questioned the desirability of this statement and were particularly anxious to know what was implied by the word 'unit'. It was accepted that in the planning of a series of lessons in mathematics aims and objectives should be both clear and explicit. But for an individual isolated lesson it was important to retain the freedom of the teacher to be adaptable and flexible. Delegates were reminded of the discussion on this matter in Professor Wrigley's lead paper where the point is made that good teachers do not necessarily work in a completely logical manner; they do not always proceed from aims and objectives to consequent behavioural changes and their subsequent evaluation.

59. It was felt by some delegates that the first statement in Paragraph 32 concerning the use of Piaget-type checks in the teaching of mathematics was rather strong. It was pointed out that the Geneva school of Psychology had its critics as well as its supporters. The general feeling of the delegates was perhaps represented by the later part of the paragraph. The use of Piaget-type testing in the Nuffield Mathematics Project had been a bold experiment, was still controversial, but the method had certainly proved useful. Such methods, if used in other situations and in other

countries, should certainly be introduced gradually, with caution, and with a full understanding by teachers which would probably only come with in-service training. Otherwise there would be many dangers.

60. There was an interesting and important discussion on Selection. It was first noted that in designing selection procedures those responsible should have clear objectives in mind when they are devising content and method. It was further felt that the phrase in Paragraph 38 "There is no doubt that such examinations have considerable influence upon the teaching in the schools" was an under-statement. Delegates were reminded of the discussion which followed the lead paper when in addition to the usual concepts of reliability and validity a new idea was introduced by the word *beneficence.* By this was meant all the possible good and bad effects of an extraneous nature which follow from any important examination. The backwash effects on teaching in schools which follow from a selection examination should never be ignored. As a result it is important to include questions of intrinsic mathematical value in the selection test.

61. There was some confusion as to what was meant in the report by a verbal test and the situation was clarified by a delegate from Trinidad. He pointed out that selection material in his country had been imported from Great Britain. Among the tests used was a verbal reasoning test. Psychologically this had been attacked because it was felt that the language of tests constructed and standardised in one country might not be suited to the culture pattern of another country. So the report meant that the verbal test (written not oral) should be one which was suited, not to the country in which it was manufactured and standardised but to the country in which it was to be administered.

62. An interesting discussion developed around the recommendation in Paragraph 46 that "ideally examinations should not be of one form" and in Paragraph 55 that tests "should include multiple-choice, multi-facet and traditional types of items". These recommendations had been made because the working group had recognised the importance of first deciding on the *function* of the test before deciding on the *type* of test. In some situations of a qualifying kind items and questions which could be answered by nearly all candidates were appropriate. But in highly selective situations where discrimination amongst the best candidates was needed then difficult items were needed. Considerations of beneficence and good backwash effects also suggested that a variety of forms of test were needed so that the teaching would not become stereotyped. These arguments were not accepted by all delegates in the plenary session. In particular the implied criticism of the possible over-use of multiple choice tests was not accepted by a Canadian delegate who said "There are those in my country who would strongly disagree with that statement and among those I would number myself".