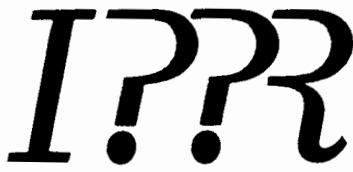141

IPPR

INSTITUTE FOR PUBLIC POLICY RESEARCH

Education and Training Paper No. 5

Assessment in Schools:

An Alternative Framework

Harvey Goldstein

IPPR

INSTITUTE FOR PUBLIC POLICY RESEARCH

30/32 Southampton Street, London WC2E 7RA
Tel: 071-379 9400   Fax: 071-497 0373

The Institute for Public Policy Research is an independent charity whose purpose is to contribute to public understanding of social, economic and political questions through research, discussion and publication. It was established in 1988 by leading figures in the academic, business and trade union community to provide an alternative to the free market think-tanks.

IPPR's agenda reflects the challenges facing Britain: changes in the European continent and the European Community; the threats posed by uncontrolled industrial growth to the world's ecological balance: the impact of new technologies and scientific developments; changing social and economic relationships between women and men; the need to decentralise power to the regions and nations of Britain.

Besides its programme of publication, IPPR also provides a forum for political and trade union leaders. academic experts and those from business, finance, government and the media, bringing people together from throughout Britain and the European Community to discuss issues of common concern.

CONTENTS

Summary

This paper argues that the assessment system introduced by the 1988 Education Act is based upon a misdirected notion of the proper role of assessment in schools. The report points out that the main assessment aim of the 1988 Act is to rank schools and teachers using the average test scores of their students, and the report argues that this is misleading and unfair. It misleads because the appropriate way to judge schools using the achievements of their students is by how much *progress* those students make during the period they are at their schools. This means that students' achievements at the time they enter school must be taken into account. It is unfair because, without being able to measure progress, it will appear to favour those schools with high achieving intakes. Furthermore, if the reported results were to be taken seriously, it would tend to stigmatise and polarise schools within a set of self fulfilling prophecies.

The paper sets out four functions of assessment: certification, learning promotion, understanding the factors associated with effective learning, and provisional screening of schools for unexpectedly high or low results. It discusses the proper role of each of these functions and argues that they need to be considered separately. The report places particular emphasis on assessment for the promotion of learning and discusses how formative and diagnostic assessment can be a positive factor in promoting the involvement of students and parents in the process of learning.

Certain implications flow from the views in this paper.

First, the public reporting of national assessments, including public examination results, with the aim of producing league tables or rankings of schools, should be stopped.

Secondly, assessment can be used to inform and benefit both policy and people. Policy needs can be served by setting up a carefully planned national system of assessing performance which seeks to understand the factors which promote learning in general. One model for such a system is the Assessment of Performance Unit, which should be revived and reviewed for this purpose.

People such as parents, teachers and children themselves can benefit from assessment when it is designed to promote a formative evaluation of individual students as part of a positive collaboration between the students, the teachers and the parents. Because it is formative, such an evaluation essentially is a private matter for those involved, enabling a student's progress to be judged both in the context of the school and also in terms of wider expectations for students in general. Existing Records of Achievement should be developed in pursuit of this aim.

Finally, it is important that a concern with assessment is not allowed to dominate curriculum organisation and content. Except for end of compulsory schooling examinations, assessment should be seen as formative and not for the purposes of grading and selection. Some influence of end of schooling examinations on curriculum content and organisation is inevitable, but schemes such as those for graded assessment linked to examinations, serve to extend the constraining effects of such examinations throughout the whole period of secondary education, and they should be dropped. End of schooling exams themselves should be considered as a part of an integrated system of qualifications within the total 14-19 educational provision of academic and vocational courses.

Since the 1987 general election the Conservative Government has made a massive investment in a programme of assessment designed to affect the whole conduct of teaching and learning in schools. The clearest statement of this programme is found in the report of the Task Group on Assessment and Testing (TGAT) which reported in 1988. With some modifications because of practical difficulties, it is the recommendations of that report which the Government has attempted, with varying success, to implement and has incorporated in the 1988 Education Reform Act. Its main features are these.

Standard assessment tasks (SATs)

There is to be a series of centrally devised assessments to be administered annually to all 7, 11, and 14 year olds (key stages 1, 2 and 3) in public sector schools. The first, non reported, implementation is planned for the summer term 1991 with seven-year-olds. These assessments will be administered and marked by teachers. The Government's claim is that they will reflect 'good practice' and one implication of this is that they will be relatively time consuming.

The original intention was to introduce these for all areas of the national curriculum but it now seems that only the three core subjects, and possibly technology will be covered. At the age of sixteen it appears that the existing GCSE will remain the predominant means of assessment.

Teacher assessment

In addition to the centrally devised tasks teachers will also assess their students, prior to the administration of the SATs. They will grade their students using the same sets of marking categories in the same curriculum areas as the SATs.* When the SAT results are available for the same students, any overall discrepancies between teacher assessment and SAT results will be removed, generally by reviewing the teacher assessment. The intention is to bring the teacher assessment into line with the SAT assessment, an implicit teacher training activity.

Statements of attainment and attainment targets

Each area of the curriculum is divided into a large number

---

* At key stage 1, as a result of pilot work in 1990, the Government has decided that the SATs will cover only a minority of the 'attainment targets'. The teacher assessments, however, will continue to cover all attainment targets.

of 'attainment targets' (ATs) each of which is intended to describe an aspect of what a student knows or understands. A statement of attainment (SOA) is a description attached to a specific task at which a student is deemed to have succeeded, and each one is linked to one or more attainment targets. Judgement of whether a student has achieved an AT is on the basis of a collection of SOAs. The attainment targets are grouped into 'profile components' which are intended to represent meaningful subdivisions of a curriculum subject. Thus, for example, in mathematics one profile component is labelled 'practical applications of mathematics'. An example of a SOA is the following. It is linked to level 2 in English, that is appropriate to key stage 1 (KS1):

> 'Spell correctly, in the course of their own writing, simple monosyllabic words they use regularly which observe common patterns.'

An example of an AT is the following level 2 in the mathematics profile component 'estimation and approximation in number'

> 'Make a reasonable estimate of a number of objects up to 30'

Statements of attainment and attainment targets usually have examples attached to them in the various Department of Education and Science (DES), National Curriculum Council (NCC) and Schools Examination and Assessment Council (SEAC) documents which describe them (see *The SEAC Guides to Teacher Assessment* issued in 1990).

Reporting results of assessments

There are two principal aspects of reporting. The first is to do with the detail presented. For example, whether individual profile components are reported or just average attainment for a subject. This issue is concerned with reporting results for individual students, to their parents etc and with reporting results on a school by school basis.

The second aspect is to do with the way in which a school's assessment results are to be presented. The TGAT report recommended that the average results for each subject, or attainment target etc. should be published along with some kind of statement about the circumstances of the school. The intention is that this will allow people to compare the 'effectiveness' of schools in different areas.

We now turn to the intentions of the Government, both political and educational.

The Control of Education

The assessment reforms have to be viewed as part of a consistent policy to centralise control of education, and in particular to remove as much as possible from the control of Local Education Authorities (LEAs). The 1988 Act sought to do the latter by the delegation of management to schools, open enrollment and the provision for schools to 'opt out' of LEA control.

In addition, the National Curriculum and National Assessment constitute a powerful infrastructure for central control. They prescribe a series of hurdles for individual students and a set of easily used comparisons between schools. Teachers and schools can no longer decide for themselves (unless of course they are in the private sector to which the National Curriculum does not apply) the broad order in which they will teach their subjects, and their professional futures may depend crucially on the test results produced by their students. Children, from the beginnings of their schooling, will be initiated into a competitive sequence of tasks to be completed before they are allowed to 'progress'. In the proposals which we set out below there is a deliberate attempt to weaken this central control by returning much of the responsibility for the administration and use of assessment to where it belongs, in the schools and the LEAs.

Along with this structure for control there is the intention to force as much as possible of a 'free market' economy onto education. The publication of each school's average results will produce 'league tables' from which parents are expected to choose the 'best' and administrators can take action on the 'worst'. That this is both invalid and irresponsible has been shown by research (Woodhouse and Goldstein: 1988, Nuttall *et al*: 1989) and we elaborate on this in a later section.

Legitimation

While it had a clear set of political intentions, the Government still needed to enlist support from key elements of the educational establishment in order to carry out its changes within a short timescale. Thus the Task Group on Assessment and Testing, the Schools Examinations and Assessment Council and the National Curriculum Council adopted a few key viewpoints. Principal among these are those of 'criterion referenced testing', 'educational progression' and that the publication of school results is a proper way to make schools accountable. In addition it has been assumed that the same system of assessment will provide both a 'formative' or 'diagnostic' tool for monitoring the learning of individual students and a suitable tool for publicly comparing the performance of schools. We shall have more to say on this point when we outline our alternatives.

By the summer of 1990 the haste with which the Government had established the new structures yielded serious implementation problems. Ministers have publicly admitted that they had imposed too large a burden upon teachers and students and have sought ways to trim what they consider to be the less 'essential' elements. The original elaborate structures which promised 'user friendly' testing have undergone considerable modification. Yet the underlying rationale for the changes has remained and continues to provide a language for supporting and developing the original patterns of curriculum and assessment.

The issue of publication of school results is now being questioned fairly extensively, but the other issues are not. The 'progression' argument is exemplified in the establishment of a series of hierarchical stages or levels within each attainment target and the criterion referencing case finds expression in the 'mastery' nature of the statements of attainment. While these viewpoints are highly questionable, it is of great concern that they have been subjected to little close scrutiny. There is not the space here to go into fine detail, but a brief critique will help to introduce the discussion of alternatives.

Criterion referencing and educational progression

The idea of criterion referenced assessment originated in the 1960s as an attempt to link assessment to learning objectives. In the 1980s it has seen a resurgence in the UK; in grade criteria for GCSE, in the graded assessment movement, in some of the early work on profiling and now in the attainment targets for the National Curriculum. Often crudely interpreted in terms of 'can do' statements, it is promoted as a provider of practical information about what a pupil has 'learnt and mastered' (TGAT para 94). That report claims that 'Norm referenced approaches conceal changes in national standards . . . Only by criterion referencing can standards be monitored." (para 222). The report, however, provides no indication how such 'standards' are to be derived and communicated.

Attempts to produce descriptions of 'mastery' based on criterion referenced ideas need to operate at a level of generality which demands a set of 'context free' descriptions. Nevertheless, the information upon which any such description is based will always be limited to the tasks chosen for assessment, and this is especially so in the case of the national assessment tasks. Consider for example the

attainment target (level 2 mathematics) 'make a reasonable estimate of a number of objects up to 30'. One would want to know whether a child could always do this, do it with small as well as large objects, when the objects are arranged in one pattern rather than another etc, etc.

To make a decontextualised statement of achievement is not feasible because we have to assume that such a statement can be applied in the far greater number of contexts which were not observed. Indeed, we know that in general it cannot be done. The work of the Assessment of Performance unit (APU) in mathematics, for example, has shown how something as simple as a change in presentation format can change performance markedly and the same is true in language assessment.

While giving the impression of being an exciting new approach to assessment, and in particular of avoiding the need to 'discriminate' between students, the appeal to criterion referencing contains much rhetoric but little real substance. In particular it claims to avoid the iniquity of 'norm referencing' which 'discriminates' between children by finely ranking them on the basis of test scores. Yet, in practice, the TGAT system can easily be used to rank students, albeit perhaps somewhat crudely on the basis of 10 attainment target levels.

In practice, recent concern with criterion referencing has helped to focus attention on the *content* of assessments rather than on their ability to discriminate between students. This has been a healthy development, but criterion referencing *per se* is no panacea. When stripped of the jargon which surrounds it, the new system remains a clear attempt to introduce ranking and grading of children across major areas of the curriculum throughout their school career.

The other major notion, that of educational progression, is again closely linked to a desire to grade students by their achievements. A simple set of levels which everyone has to progress through in the same order provides a neat way to classify students. Yet the educational evidence that this is either necessary or desirable is extremely thin (see Noss *et al*: 1989 for a discussion of this point in the context of mathematics). Learning and teaching sequences need not be identical for all children, and any attempt to make them so goes against the best principles of teaching and learning and discriminates against those students who cannot conform easily to the laid down pattern. The further linking of higher levels to GCSE effectively brings the certification process to younger age groups than previously, with the attendant concerns about streaming and overconcentration on obtaining high test scores.

3. A RATIONAL AND JUST APPROACH TO ASSESSMENT

In setting out an alternative framework for assessment policy we have chosen to concentrate on the ways in which assessments will be used rather than on the details of their construction or content. The latter, of course, are important and one of the challenges facing the adoption of an alternative programme is to design assessments consonant with the uses we shall identify. The uses themselves come first and need to be related to a coherent educational and social philosophy.

A fundamental tenet of that philosophy is the stress on education as a socially co-operative enterprise rather than a marketplace of competition. We do not deny competition, and recognise it as a motivating force at individual and school level, as in public examinations and team sports. Rather, we believe that it is possible to develop an assessment policy which also promotes social ends; which obtains information to improve the overall quality of education without in the process condemning the already disadvantaged. It is pointless to deny that children differ in their achievements and that high achievement, in whatever sphere, should be recognised. We do believe, however, that a just system should be as concerned with those who learn slowly as with those who learn fast.

We believe that one of the strengths of any system lies in its variety of approaches and we would seek to encourage and support experimentation with both curriculum and assessment. For this reason we do not deal with detailed issues such as whether or not to have mixed-ability grouping or whether to abolish age-grouping. We envisage that these are matters for continuing debate and wish to make no assumptions about the universal relevance of any particular method of organisation or management. Above all, we recognise that on many issues we are still ignorant and the tone of our proposals is that of pursuing knowledge for improvement within a belief that this entails a cooperative effort based upon mutual trust rather than mutually destructive competition.

In the following sections we apply these approaches to forming a critique of present Conservative Government policy, and to the development of just alternatives.

The uses and functions of assessment

We start by distinguishing four major uses of educational assessment which should determine the form of an assessment system.

● As certification, to provide entry to employment or a further stage of education.

● To promote learning by establishing what a student's strengths and weaknesses may be.

● To understand those factors which promote learning.

● As a screening instrument to identify schools or departments which may be poor promoters of learning.

Our argument is that these four functions serve distinct purposes and require distinct approaches to their introduction. An integrated assessment policy will need to address the relative importance of these functions and the links between them.

Certification

The role of external examinations is to provide a standardised individual selection mechanism which enjoys the confidence of its users. At present the GCSE provides the principal examination at the end of compulsory schooling and affects the structure of the curriculum at least from the age of 14. There is a clear need to integrate the GCSE within a total 14-19 provision of courses and qualifications and a discussion of this is given in the IPPR report *A British 'Baccalauréat'* (Finegold et al: 1990).

Changes in certification, especially in the GCSE, should be seen as developmental, building on what currently is available. In contrast to the implementation of the 1988 Act, the changes introduced with the GCSE have been the subject of debate and experimentation for over a decade. We see as particularly important the introduction of coursework assessment and the attempt to remove the divisiveness which characterised the old separate O level and CSE exams.

Whatever new certification procedures emerge, their influence on the early years of education should be minimised. It has long been recognised that there are negative consequence of tests or examinations which have 'high stakes', that is their results really matter to those involved. These consequences have to do with the narrowing of the teaching effort and the process of learning in order to concentrate on maximising test or exam grades. We consider it educationally desirable to postpone such effects till as late as possible during a student's school career.

It also follows that any scheme which seeks to extend to younger ages the certification system should be avoided. For this reason we are opposed to those schemes of graded assessment, such as in mathematics and science, which envisage the award of certificates upon achievement of predetermined levels or stages. Since such achievements can occur, in principle, any time during the secondary years such schemes effectively encourage a concern with certification throughout secondary schooling. The same intention is apparent in the linking of national assessment levels to GCSE gradings, and in attempts to force the GCSE to reflect national curriculum attainment targets.

Assessment for the direct promotion of learning

The use of assessment to promote individual learning is quite different from the other uses set out above. This 'diagnostic' function needs to probe each student in depth to expose weaknesses as well as strengths, over a wide range of activities. As such it has to encourage openness and honesty so that these can be appraised and acted upon. It is inconsistent with certification which encourages students to minimise their weaknesses and exaggerate their strengths. It is also inconsistent with any system which is used to judge schools or teachers, for similar reasons. Since the present system of national assessment is primarily one whose intention is to provide such comparative judgements, it cannot legitimately be considered as adequate for any kind of diagnostic function. The effective omission of this function from the present national assessment system is serious because that function is likely to be downgraded in importance.

The use of formative or diagnostic assessment by teachers is, of course, not new. Yet it is apparent that the training of teachers in this activity is generally poor and is unlikely to be enhanced under the existing system. In recent years, however, a suitable vehicle for this kind of assessment has emerged, namely the Record of Achievement (ROA) and the development of these should be encouraged.

A great deal of work has already been carried out and this should be built upon. The present Government has given them a lesser importance than the national assessments and they will not flourish unless encouraged actively with supporting resources. This requires pre-service and, particularly, in-service training and the provision of time for teachers to cope with it.

A particular aspect of ROA which should be encouraged is the formative one where students and teachers collaborate on assessment and the student takes some responsibility for his or her own assessment. The link between such formative records of achievement and a 'summative' document which is for external consumption as an end of schooling report should be weakened. Linking a formative, essentially private record, with one which is for presentation to others, weakens the diagnostic function of the former. Instead, schools should be encouraged to set a series of common but 'locally referenced' targets of achievement. These might relate to intellectual, physical or social achievement not formally part of an external accreditation or exam system; for example athletic, leadership activities, performing artistic activities, chess, swimming etc.

By locally referenced we mean that these achievement targets are those actually attainable by the students in a particular school. Because the assessments are formative

they are best described within the relevant local context. Typically this is the student's classroom or school, but for some students such descriptions may range more widely, for example in competitive sporting activities. The aim of these targets would be to give proper recognition to a wide range of legitimate and important activities without linking them to a summative report.

We should make it clear that locally referenced assessments are also compatible with wider national norms. With proper training and support, and where it is useful to do so, teachers can relate these assessments to expectations of what students in general are capable of achieving at each stage of their school career. This can then be communicated both to parents and their children.

An important aspect of formative assessment is that it measures progress. The progress made by a student, from a given starting point, is amenable to teaching. By contrast, the student's achievement at any particular time is only partly determined by what happens in school: it is also, and perhaps predominantly, affected by environmental, social and biological factors over which schools have little control. By concentrating on progress, and shifting emphasis away from assessment at a single occasion, we therefore move towards what legitimately can be expected from schools and teachers. We shall discuss this issue again later.

It is important to find ways in which schools can work actively with parents so that they can gain better understandings of their children's learning. The involvement of parents in the preparation of records of achievement offers considerable potential for this. It becomes a valuable means of strengthening parental involvement in the school and with the education of their children, and we should give a high priority to encouraging such activities.

Assessment for monitoring

We discuss below why external assessment systems are unsuitable for comparing schools. Nevertheless external assessment can help our understanding of the factors which promote learning and long term research on this needs to be encouraged. A useful model here is the Assessment of Performance Unit (APU). On a sample basis the functioning of a varied curriculum can be monitored and analysed in conjunction with the LEA's. Assessment tasks would be developed, possibly based upon some of what is already available, and using the accumulated experience of the APU to regularly analyse, nationally and locally, the results and to initiate debate on those results.

Because they would be conducted on a sample only, they would not need to impose a large burden on schools or students. Likewise the results should *not* be used to label or

rank schools or LEA's. They would, however, be one means by which the impact of different approaches to implementing a restructured national curriculum can be studied. These surveys would collect information about a wide range of academic and non academic activities and the circumstances in which they occur. The achievements of pupils would be studied in relation to the contexts in which schooling occurs; in particular, the way in which changes in curriculum, resources, etc. are seen to affect those achievements. More than any other assessment activity, these assessments would provide a systematic means for judging the effects of educational innovation and change. If the notion of 'monitoring standards' is to have any useful content, this activity would be an important means of providing it. It would underline a philosophy of cooperative enterprise in the sense that the lessons learnt from it will be used for the benefit of all.

Suitable institutional arrangements for carrying out these surveys should be developed and they should not be controlled by government departments. The confidentiality, impartiality and integrity of such monitoring should be safeguarded explicitly, so that there is an onus to publish findings and to have a wide discussion of procedures.

Assessment for Screening

The research evidence suggests that the earlier achievement of students is the most powerful predictor of their later achievement, with socio-economic factors being important also. Unfortunately, we lack very good evidence about the influence of school contextual factors such as school organisation or interpersonal relationships within a school.

It is clear, however, that the proposals of the present Government for reporting school assessment results are unfair because these 'raw' results will reflect largely the achievements of the students when they enter a school as well as social factors. Furthermore, there is an inherent instability in attempts to rank schools based upon average school results, even when it is possible to take account of socio-economic or intake achievement scores (Woodhouse and Goldstein: 1988). Any attempt to do so, whether by a formal analysis or an informal procedure such as suggested by TGAT, will lead only to lengthy and inconclusive debate which will throw little light upon the genuine reasons for any school differences. In particular it may well ignore genuine excellence and overlook genuine failure in a quest for simple league tables.

It is now generally agreed by most researchers that comparisons between schools should be based upon analyses of individual student data rather than school averages. The use of so called 'multilevel' models in analyses of

achievement has pointed to the depth of information which it is possible to extract. Nuttall et al (1989) analysed examination data from the Inner London Education Authority and showed that schools differed along several dimensions. They differed in terms of the achievements of different ethnic groups, in terms of the initial intake of their students and in terms of the gender difference in examination achievement.

It is clear that a prerequisite for a fair comparison of schools is to be able to take account of intake achievement and to analyse individual student data. While these basic requirements for useful school comparisons are recognised, we are still a long way from being able to prescribe a standard analysis which can be adopted routinely to provide definitive school comparisons. Rather, such analyses as are possible must be regarded as research tools. Statements about individual schools will be tentative, and the whole exercise should be regarded as a screening procedure which could indicate which schools might repay further study. Any exercise of this kind would be carried out in cooperation with schools, inspectors and advisers and in a spirit of cooperation and support rather than one of public judgement.

Within our overall concern to promote useful and relevant assessment, this one does not have a high priority. Nevertheless, we are persuaded that research in this area should be promoted actively, building upon that which has already been done, and linked with that on assessment for monitoring.

4. MONITORING THE IMPLEMENTATION OF THE CURRICULUM

While we have been concerned with assessment, a few final words can be said about the need to monitor the implementation and functioning of any new or modified curriculum.

We do believe that it is both possible and desirable to make valid judgements about important aspects of the performance of schools, but we have argued that the use of student assessment results to do so is wrong. Implicitly, the use of assessment results to monitor the curriculum and its implementation is mistaken. If we wish to understand the curriculum and how it functions then that should be done directly rather than indirectly via student responses to it. A system of within-school review with external moderation and advice is possible, and indeed desirable. This would allow debate and change to occur, without the implicitly threatening implications of external testing. It would be an essentially local activity with some national coordination, and would involve parents and governors as well as schools and LEAs. In other words, if we wish to see how well the curriculum is working we should directly observe its operation, rather than indirectly trying to infer its effects by measuring the achievements of its students.

References

Dept. of Education and Science (1988) 'National Curriculum: Task Group on Assessment and Testing, A report'.

Noss,R, Goldstein,H, and Hoyles,C (1989) 'Graded assessment and learning hierarchies in mathematics'. *British Educational Research Journal*, 15, 109-120

Nuttall,DL, Goldstein,H, Prosser,R and Rasbash,J (1989) 'Differential school effectiveness'. *International (1989): Journal of Educational Research*, 13, 769-776.

Woodhouse, G. and Goldstein, H. 'Educational Performance (1988):Indicators and LEA league tables'. *Oxford Review of Education*, 14, 301-320.

Finegold,D, Keep, E, Miliband,D, Raffe,D, Spours,K, and Young,M (1990) *A British 'Baccalaureat'.* London; Institute for Public Policy Research.

Education and Training Paper No. 5

The Government's plans for the testing and assessment of 7, 11, and 14 year olds are deeply controversial. This report examines the facts and rationale behind the Government's case, and then suggests an alternative course.

The report argues that a just and successful system must be as concerned with those who learn slowly as those who learn fast, and details the functions that an assessment system must serve if it is to promote quality teaching and learning.

Harvey Goldstein is Professor of statistics at the Institute of Education, University of London. He is a specialist in educational research methods.

I??R

£5.00