

## REPORTS

## A curriculum for teacher assessment

CAROLINE GIPPS and HARVEY GOLDSTEIN

**Introduction**

The British government is moving rapidly towards the fulfilment of its plans for the systematic testing of all 7, 11, 14 and 16 year-olds. One of the biggest and most far-reaching educational innovations has been conceived, nurtured and come to maturity in a remarkably short time. Its conception occurred during the 1987 general election campaign. It has been nurtured within the Task Group on Assessment and Testing (TGAT) (DES 1988 a, 1988 b) and, following the announcement of a substantial dowry, it has attracted suitors nationwide (*The Times Educational Supplement*, 15 October 1988). At the time of writing, we know which three agencies are to develop Standard Assessment Tasks for seven-year-olds. We also know that there has already been criticism of the proposed tasks (*The Sunday Times*, 18 December 1988, *Guardian*, 3 January 1989).

In this article we shall examine the recommendations of the TGAT report, and in particular the sections on teacher assessment. Much of the (surprisingly warm) reception given to this report dwelt on the supposedly 'progressive' proposals for involving teachers in the assessment and testing. Further reflections and a detailed study of the actual recommendations, however, have led us to a quite different view, which we shall now elaborate.

**The role of teachers**

There are two distinct kinds of assessment discussed in the TGAT report. One is a series of centrally designed 'standard assessment tasks', both written and practical. The report spends time arguing in favour of 'innovative' and interesting tasks which can be incorporated into daily teaching. At the moment there seems to be a commitment by

the Department of Education and Science (DES) towards innovatory kinds of tasks such as those pioneered by the APU and endorsed by the TGAT report. These tasks will be marked by teachers who will receive relevant training.

The other kind of assessment is that to be done by the teachers themselves on the basis of their pupils' general work, in the same 'profile component' areas covered by the centralized assessment. The report devotes much space to describing how the teachers' results are to be made compatible with each other and with the centralized assessment. The report recommends that 'teachers' ratings be moderated in such a way as to convey and to inform national standards' (DES 1988 a: para 62). It suggests that, if left alone, 'teachers' expectations [of what is normal] become the teachers' standards' (ibid.: para 65). The report recognizes that 'teachers' rank orders... may vary systematically from rank orders provided by test users (ibid.: para 66)', and so the notion of teacher assessment adopted by the TGAT is one where such differences are eliminated.

But, we would argue, where teacher assessment is to be used for discussion, negotiation and recording with pupils, teachers and parents participating, then the prime requirement is not to convey national standards. Furthermore, such locally based assessment is in many respects more appropriate as the basis for decisions about curriculum provision, individualized teaching schemes and so forth. It is precisely its ability to reflect local conditions which makes it valuable. It is only where *comparability* is paramount that the requirement to relate to national norms is necessary. Yet neither the first report nor the supplementary reports recognize this distinction and by implication, therefore, would seem to place lower value on those

---

*Harvey Goldstein* is Professor of Statistical Method and *Caroline Gipps* is Lecturer in Curriculum Studies at the University of London Institute of Education. They have worked together on a number of research projects and publications on assessment.

elements of teacher assessment which do not accord with the centralized assessment. A likely consequence is that teacher assessment could become restricted to just those things which can also be measured by centralized tasks.

The TGAT report proposes that groups of teachers should meet, with samples of their students' work, and after suitable discussions should form a consensus about grading standards and adjust their overall grades so that the proportion of their pupils achieving each 'level' agrees with the proportions achieving each level on the centralized tasks.<sup>1</sup> An immediate consequence is that, in terms of reporting school (or LEA) results, the teacher assessment is strictly redundant because it has been forced to agree with the centralized assessment results for the school. There is therefore little point in forming a 'combination of moderated teachers' ratings and standardized assessment tasks' as the TGAT recommends (DES 1988 a: para 63). It is also worth pointing out that if teachers' assessments were required to agree with the centralized assessment, say at the local education authority (LEA) rather than the school level, then a similar point would apply but just to LEA comparisons. It is quite possible, however, that the moderated teacher assessment component will come to be seen as expendable: the government is concerned at the cost of the proposed moderation exercises (DES 1988 c).

#### **Standards: local or national?**

We have already touched upon the distinction between locally (i.e. within-school) relevant assessments and nationally comparable assessments. It is fairly clear that any system of assessments which is intended to provide comparisons between units of an educational system must be based upon a common currency. Thus 16+ examinations have such a currency, expressed in grades. Of course, the equivalence between one exam group's grade A and another group's grade A is as much a matter of fiat as of scientific observation; but what counts is society's acceptance of the coinage.

Within the plans of the present government, designed to encourage a free-market economy in education, the currency unit is a test score.<sup>2</sup> The 'worth' of a school

is intended to be measured, to a large extent, by its average scores in different curriculum areas. These averages may be expressed in terms of, say, the proportions achieving each 'grade' or, more simply, as a single mean grade.

Just as the exam groups have had an elaborate mechanism buttressing their claims for grade comparability, so the national testing will have its framework of attainment targets, common tasks, moderation and so forth. In fact, the problems facing national assessments in terms of equivalence, bias and fairness are at least as severe as those faced by the exam boards. We shall not dwell on these here, but see TGAT (1988 a: appendix F) and Goldstein and Cuttance (1988).

Our point is that assessment has always played a key role in children's education at all ages. This assessment is typically embedded within the very process of teaching and learning, and has little to do with national norms. Rather, it seeks a detailed evaluation of learning against specific curricular criteria. It often addresses what parents want to know about their children, for example how they are coping with language problems or design technology, or in their physical development. This form of assessment has to do with individual progress and at its best is established upon a basis of partnership between student, teacher and parent. Its most explicit recent formulation is found in the records of achievement projects (Broadfoot *et al.* 1988), and the ideal conditions for it to flourish seem to involve the *absence* of a pervasive national assessment scheme.

Within such 'local' assessment we can still compare children and schools. Local comparisons of achievement, whether in swimming, science or chess, can be both illuminating and motivating. Precisely *because* they are local, and hence within an understandable context, they can acquire a meaning which is denied to those assessments which are required to have national comparability.

The imposition of national assessment presents a real danger to this form of professional assessment. Yet this does not need to happen, and in our final section we outline a positive programme to support professional assessment. This would aim, through a programme of teacher education, to create an awareness of assessment and an appreciation of the expanded opportunities

offered by an imaginative interpretation of the government's proposals which is grounded in sound pedagogical practice.

### **A curriculum for assessment**

If we cannot avoid the comparability element in national assessment, it is important to emphasize where the scope lies for the positive, professional use of assessment within this structure. In effect we wish to propose a *curriculum for assessment*.

In terms of first principles the main beneficiary of assessment must be the individual child, and therefore the main purposes must be diagnosis and motivation. The latter has been important in the development of many new approaches to assessment: graded tests, records of achievement, and aspects of GCSE. Motivation, however, is little addressed in TGAT, although the underlying model (not explicitly acknowledged) is the graded assessment one. Competition will, of course, provide motivation—especially for those who are already successful.

One lesson of the records of achievement developments (Broadfoot *et al.* 1988) has been that where teachers and students discuss progress and assessments in a fair and open way, then significant changes can take place in student motivation, confidence and attitude to learning. It is vital, therefore, that there be room for this process within the new assessment arrangements. We know, however, that teachers are not particularly good at this, their natural tendency being to dominate in discussions with students. Training and awareness-raising are therefore important here and schools which have taken part in developing their own or the LEA's records of achievement may well find that they have a head start.

Another way to enhance teachers' professional role in national assessment would be to use the SATS as a testing and learning device. By observing children carrying out assessment tasks (as the APU has done with diagnostic interviews), by prompting where the child gets stuck or goes wrong, teachers can really begin to assess what children know, understand and can do, and this diagnostic information can inform teaching.

This does, however, raise problems for the 'objective' reporting of achievement. Traditional theories of assessment tend to

be based upon the view that achievement is just a characteristic of the student which, in principle, can be measured or rated. These are essentially all developments of the 'empty vessels' view of learning whereby one simply needs to find a good way of measuring how much has been poured in! From such a perspective, 'helping' a student during assessment would be associated with 'cheating'. A more dynamic view involves teachers as active participants in a process that does not separate assessment from the process of learning. Necessarily this view will rely more heavily on teacher judgement and hence on adequate training.

This kind of assessment has typically been the province of the educational psychologist (or advisory teacher) in the past. In our view the necessary skills should become a part of every teacher's repertoire. They are not so complex that they need to be restricted to one professional group and the SAT may well be the peg upon which to hang this new hat.

It is clear that there will be a programme of teacher education, enabling teachers to administer and mark the SATS. This will be important, because the tasks need to be administered in a standard, common way if national comparability is to be credible; also, if the tasks are to be innovative and broadening, then they will be quite new to most teachers. More importantly, however, we believe, teachers should become competent assessors themselves and should have an enhanced role, such as we have described, within any national assessment framework. This too must be part of the overall programme of teacher education.

What we have done is to outline not only a curriculum for assessment, but a curriculum for teacher training in assessment. Too many teachers are assessment-illiterate when it comes to formal assessment. This is not their fault; initial and in-service training has too often ignored assessment. In addition, the difficulty of finding out about and obtaining standardized tests and the mystique of mental measurement have militated against teachers acquiring the appropriate skills.

Assessment *is* a tool for teachers, to be used for the benefit of students. It should be developed so that the obverse of competition and global comparisons is one of feedback, enhanced awareness and motivation. This is sound pedagogical practice, and it is also one way of enhancing

(some may say restoring) the professional role of teachers.

Finally, little of what we are suggesting can be achieved by adding it to the already burdensome requirements imposed by recent legislation. It will be for individual LEAs, schools and teachers to set our view of assessment alongside their other responsibilities and to assign it the priority they feel it deserves. We merely aim to point out that there is a broader view than has so far been offered and to sketch it out.

#### Notes

1. In the first supplementary report it is suggested that this adjustment or 'reconciliation' process should also involve the proportions achieving each level nationally on the previous year's assessment. No details are given, and it seems curious that a difference between a school's distribution and the national distribution, should need reconciling!

2. We use the term 'score' to refer to any quantification of assessment results, whether this be a number as in a standardized test or a grade as in public examinations.

#### References

- BROADFOOT, P., JAMES, M., McMEEKING, S., NUTTALL, D. and STIERER, B. (1988) *Records of Achievement. Report of the National Evaluation Schemes* (London: Her Majesty's Stationery Office).
- DES (1988 a) *National Curriculum: Task Group on Assessment and Testing* (London: Department of Education and Science).
- DES (1988 b) *National Curriculum: Task Group on Assessment and Testing: Three Supplementary Reports* (London: Department of Education and Science).
- DES (1988 c) Kenneth Baker sets out principles for assessment and testing in schools. Press release.
- GOLDSTEIN, H. and CUTTANCE, P. (1988) National assessment and school comparisons. *Journal of Educational Policy*, 3 (2): 197–202.

## Assessment study group

BOB STAKE

At informal gatherings during the past two years an international group of educational researchers has been advocating formal study of the effects of changes in assessment policy, particularly those changes affecting the curriculum and classroom practices. Constituted of both university and government based researchers, the group has become identified with the acronym of ECAP (Effects of Changes in Assessment Policy). Open meetings have been held at New Orleans (AERA), at the Ontario Institute of Studies in Education, at Norwich (BERA) and at the University of Illinois at Urbana-Champaign, and others are planned. At the close of the initial meeting, the following manifesto was drawn up.

#### Manifesto

We the undersigned, all involved in the assessment of education, having reviewed assessment policies in schools and states and also existing research, collectively call for immediate and thorough study of the

effects of assessment on educational practice and purpose.

Assessment has been widely undertaken as a step believed necessary to improve the quality of our schools. Testing and other forms of evaluating teaching and learning are increasingly relied upon to inform professionals, policy makers, and the general public. Additionally, assessment requirements are used to arouse public concern and redirect professional effort.

Calls for technical monitoring of education are widely approved. Most people—including winners and losers in examination competition—regard testing as sufficiently accurate, fair, and useful to justify its growing emphasis. Assessment is perceived as part of responsible governance.

The need for careful governance of assessment is less apparent. Especially now, dispassionate observation and scientific analysis of assessment policy and actual practice are needed. The effectiveness of assessment as an instrument of positive change has been based largely on reason and