

# FORUM

**for the discussion of new trends in education**

Autumn 1979

Volume 22 Number 1

85p

## **This Issue**

### **The APU threat? Self-evaluation as the answer**

Joan Simon investigates the  
Assessment of Performance  
Unit

Edgar Stones and Harvey  
Goldstein expose testing  
procedures

Joan Shapiro warns of US  
experience

Maurice Galton, John Elliott  
and George Varnava present  
an alternative strategy

A R Delves and John Watts  
present Australian evidence  
for an alternative

Discussion and Reviews

going comprehensive since 1929

# Editorial Board

# Contents

Vol 22 No 1

September 1979

**Michael Armstrong**, Countesthorpe Upper School, Leicestershire.

**Clyde Chitty**, Earl Shilton Community College, Leicestershire.

**Michael Clarke**, Little Hill Junior School, Leicestershire.

**Annabelle Dixon**, Chalk Dell Infant School, Hertford.

**Margaret Gracie**, West Moors Middle School, Dorset.

**H Raymond King** Ex-Headmaster, Wandsworth School, London.

**Roger Seckington**, Earl Shilton Community College, Leicestershire.

**Byron Thomas**, Broughton Astley Primary School, Leicestershire.

**Peter Thomson**, Judgemeadow Community College, Leicestershire.

**Roy Waters**, ILEA Inspectorate.

**Harvey Wyatt**, The Woodlands School, Coventry.

## Editors

**Brian Simon**, School of Education, University of Leicester.

**Nanette Whitbread**, School of Education, Leicester Polytechnic.

**H Raymond King** From Grammar to Comprehensive (1926-1963) 4

**Joan Simon** What and Who is the APU? 7

**Edgar Stones** The World of APU 12

**Harvey Goldstein** The Mystification of Assessment 14

**Joan P Shapiro** Accountability — a contagious disease? 16

Discussion 19

**Maurice Galton** A Constructive Response to the APU 20

**John Elliott** The Case for School Self-Evaluation 23

**George Varnava** Self-assessment at a London Comprehensive 26

**AR Delves & John Watts** A Year of Evaluation 30

Reviews 30

ISSN 0046-4708

*Editorial Communications.* MSS and contributions to discussion (800 words maximum) should be addressed to the Editor, 11 Pendene Road, Leicester, LE2 3DQ. Tel: Leicester 705176.

## Business information

Correspondence relating to subscriptions etc, should be addressed to The Manager, 11 Beacon Street, Lichfield WS13 7AA. Tel Lichfield 51159.

Reductions available on bulk orders of current number. (e.g. 10 copies for £6.)

**Forum** is published three times a year, in September, January and May. £2.50 a year or 85p an issue.

## The Next Forum

Two main issues form the foci for articles in the January, 1980 **Forum**.

First, the impact of present government policies on education. Caroline Benn writes on the perspectives for comprehensive education; Mick Farley on the effect on provision for the 16 to 19s. The problems of administering local systems given the present and proposed cuts is also analysed. An overall assessment of the very serious implications of government policies will be made by the editors.

Second, a group of articles continue the focus on classroom practice. Maureen Hardy (editor of **At Classroom Level**) contributes on 'Talking in school', a study of infant school speech, and D A Reay on 'Hidden streaming', on junior school practice. The NFER 'Mixed Ability' project report will be reviewed, as well as L S Hearnshaw's remarkable book, **Cyril Burt, Psychologist**.

# Cuts and Accountability

This autumn the truth and full implications of the summer's speculations, leaks, circulars and ministerial announcements about education cuts will become known in ominous detail. Even as we go to press it is evident that Tory government plans are under way to destroy the foundations essential for the continuing evolution of that more open and humane education system for which **Forum** has campaigned over the last twenty-one years.

Teachers and parents will need to be sharply vigilant and learn to collaborate as never before to defend the education service against the massive sabotage envisaged by the most reactionary central government of the post-war era and the complicity of backwoodmen in the LEAs. For there is no doubt that 3% cuts in this and 5% cuts in the next financial year cannot be imposed without **irreparable** damage to the education of this generation of children.

A quite phoney case is being presented to the public that extensive cuts are 'necessary' and that 'the nation cannot afford' the sort of educational provision that **Forum** regards as a prerequisite for a democratic society which believes in the development of human potential. Typical of the deceit is the government's determination to include in their new Education Bill arrangements for assisted places in private schools to the tune of £50m, and a new appeals procedure whereby LEA allocation plans for comprehensive schools may be thrown into chaos. Along with the speedy repeal of the comprehensive reorganization Sections 1, 2 and 3 of the 1976 Act, these schemes expose the dogmatic ideological stance characterising the cuts. Under the banner of 'parental choice' elitism is to be subsidized and the nonsense of 'co-existing' selective and comprehensive secondary schools perpetuated and extended at the expense of quality in provision for the majority of children.

The hypocrisy of the clamour about educational standards is revealed by cuts which patently can be achieved only by such measures as further reducing nursery education, refusing entry to rising-fives, creating larger classes in primary and secondary schools, curtailing curricular options and restricting specialist teaching in many secondary schools, reducing the supply of book and other learning resources and imposing a moratorium on the in-service courses needed to enable teachers to improve their expertise. Such measures must be particularly harmful to the development of mixed ability work, to multiracial schools and to those children whose 'special educational needs' were recognised in the Warnock Report. The opportunities for improvement in the education service made available by the incidence of falling rolls, as we argued in **Forum** vol 21 no 1, are to be wantonly sacrificed on the altar of the new Tory dogma.

The context of the new government's destructive policy of educational retrenchment makes the subject of this Special Number of **Forum** even more significant than when the Editorial Board first planned it. The kind of humane education for which we stand is threatened not only by expenditure cuts but also by arbitrary normative testing and consequential curricular distortion. We have previously

warned of the misconceptions and retrogressive direction inherent in the Black Paper lobby on standards, contending that *all* children's educational achievement can be progressively raised in a genuinely comprehensive system of primary and secondary schools with flexible nonstreamed teaching groups where individual progress is monitored by teachers.

Two earlier numbers, vol 18 no3 and vol 20 no 3, were particularly concerned with assessment issues and in the editorial to vol 21 no 2 we promised 'to keep watch on the activities of the Assessment of Performance Unit and the use made of data obtained from its nationwide sample testing.' Hence we decided to present a thorough-going critique of the APU and all its works and at the same time to offer practical suggestions for an alternative strategy whereby schools may monitor their own effectiveness.

To do so is opportune following Mark Carlisle's public commitment, as Secretary of State, to require LEAs to publicize schools' examination results and the ill-omened appointment of Rhodes Boyson, Black Paperite and arch protagonist of mass normative testing, as Under-Secretary of State. Under this administration we can have no confidence in the use that may be made of the APU's alleged 'findings,' especially when local corporate management is seeking plausible grounds for cuts and school closures; and the Inspectorates' thinly veiled hostility to non-streaming, revealed in **Forum** vol 21 no 2, can be expected to be encouraged and exploited as enlarged classes make it more difficult for teachers to give individual attention. Rumour that the APU is to be exempt from cuts enhances our misgivings but causes no surprise.

The surreptitious growth of the APU under the aegis of select HMIs has been investigated for us by Joan Simon in a key article which should alert readers to dangers of irresponsible curriculum manipulation and crude devices for teacher accountability. Professor Stones' exploration of the APU world exposes many absurdities and Professor Goldstein's informative critique of the favoured assessment model's inadequacies heightens suspicion of the enterprise.

It becomes clear that the type of tests being developed for and by the APU cannot help teachers to identify children's strengths and weaknesses more accurately nor to teach better.

An American contributor, Dr Joan Shapiro, provides some warning insights into the British scene in the light of the longer experience the USA has suffered from the educational disease now being cultured here in the interest of accountability.

Four articles then offer an alternative strategy for ensuring that schools maintain and raise their educational standards by undertaking systematic self-evaluation in accordance with their own openly asserted objectives. Such an approach is consonant with professional responsibility in a democratic society and appropriate to a comprehensive system serving all children to which this journal is committed.

# From Grammar to Comprehensive (1926-1963)

---

## Raymond King

H. Raymond King has been Chairman of the Editorial Board since Forum's inception and was invited by the Board to contribute this personal, retrospective review of five decades in the evolution of comprehensive education.

---

This issue marks the 21st anniversary of the first number of **Forum**. My involvement in the comprehensive idea and steps towards its realisation antedate the foundation of **Forum** by some twenty years and are the subject and occasion of this article.

The most notable contribution to English education in the first forty years of the century was the extension of a gradually improving secondary education from the comparatively few to something approaching a quarter of the child population. At the time of my appointment to the headship of a grammar school in 1926, the Hadow Committee was on the point of presenting the Report that foreshadowed 'secondary education for all'. When this was brought about by the Act of 1944, the schools that had led the great expansion drew aside from the main stream of secondary development, to their own loss and that of the nation. The LEA grammar schools, clinging to the prestige of the independent and semi-independent sector, were mostly brought in too late, too inexperienced in the new order, and too half-hearted to give any effective lead in the comprehensive system. Comprehensive re-organisation has remained incomplete and, with strong grammar school elements still outside, has not yet won the confidence of a considerable section of the indestructible middle classes.

But by the late 'thirties some few grammar school heads had discerned and approved the shape of things to come. Secondary expansion had increased the proportion of pupils for whom academic curricula and provision were proving unsuitable. Lacking scholastic or vocational incentive, these pupils – the 'C Forms' of the typical 3-form entry school – aggravated the problem of early leaving. A few favoured schools had been able to establish engineering 'sides'. In London the junior technical schools of engineering and building offered the kind of provision and incentives that kept pupils at school for a three-year course, 13-16. Central schools and some of the senior schools were proving increasingly successful in retaining pupils for a five-year course. Hence the idea of courses with a technical or other practical bias, parallel with the academic in a 'Multilateral' school, found considerable support among secondary heads.

In London in the 'thirties a group of headmasters met regularly to discuss educational problems at greater depth and with a freedom that the meetings of their professional association did not offer. When in 1939 evacuation scattered the schools, four of these headmasters continued the discussions. In 1942 they published a pamphlet of some 15,000 words: **A Democratic Reconstruction of Education**. It argued for secondary education for all through the expansion of the curricula and provision of existing secondary

schools to meet the needs of all: in effect the first professionally conceived blue-print of what later came to be called the comprehensive school.

I have space for only one quotation to illustrate the scope and, forty years on, the relevance of the pamphlet: The normal type of school should be the day school, where influences of home and of the local community play their part equally with the school in the education of the pupil. *This partnership of home, school, and community is essential* not only for the natural development of the child, but also for the creation of a democratic society.

The circulation of the pamphlet led to the formation of a standing Conference on the Democratic Reconstruction of Education (CDRE), at the inaugural meeting of which H C Dent, editor of the **TES**, addressed a widely representative gathering. The Chairman of the LCC Education Committee was present and spoke. A Committee was elected on which the future head of the LCC secondary branch and a leading Director of Education served. Distinguished educationists gave their support.

The pamphlet's proposals included the public schools. On the publication of the Fleming Report CDRE sent its powerful and cogent counterblast to every Member of Parliament.

The campaign continued from 1942-1946 and as Chairman of CDRE I found myself in lively debate in various parts of the country, most often at what proved later to be the trouble spots.

The four headmasters, still known at Gordon Square as the Four Housemen of the Apocalypse, saw their conception of the expansion of the grammar school realised in a few schools in London. At Wandsworth, a school for 2000 boys, the LCC brought together the whole range of secondary provision and resources, the aim being to make them available to all pupils as they developed the varying abilities, bents, motivations, and vocational aspirations to take advantage of them. All roads could lead to the sixth form. But provision of this order and deployment was quite beyond the resources generally available, and comprehensives had to be defined, often incongruously, in other terms. The term 'comprehensive education' awaits definition.

Before the London schools returned from evacuation, and so while educational arrangements were still fluid, I submitted a two-tier comprehensive plan based on two contiguous schools, the one to serve as upper the other as lower school. In discussions with the Education Officer and Chief Inspector the plan was favourably considered. In the event however, the LCC scheme for grouping the future components of proposed comprehensives overtook this

idea, but gave opportunity for the submission of a second local plan. A main counter argument was that re-organisation should begin from the bottom of the scale 'upwards' and not from the grammar school 'downwards'. In the end the meeting agreed that as a move towards what was eventually intended the plan was sound and workable. However, comprehensive grouping came to nothing when it was realised that the coming 'bulge' and the inflow of large populations to the planned housing estates in our area would require all the extra accommodation of the planned comprehensives.

In 1947 the LCC established eight 'interim' comprehensives, based on central schools — ie starting from the middle — and, by exception, asked Wandsworth School to take over the London-based complement of the evacuated junior technical school of the Brixton School of Building, which had no settled home when evacuation ended.

## First comprehensive blueprint

At a meeting in 1950 of the consultative committee of heads of 'interim' comprehensives, the Education Officer handed round copies of a booklet, **The Comprehensive School**, that had been supplied to the LCC Education Committee and its Administration and Inspectorate. The booklet, of which I was the innominate author, had been published by the English New Education Fellowship with whom the CDRE had arranged joint conferences in 1944-45, and who had continued to study comprehensive re-organisation in a series of further conferences which were reported in its journal, the *New Era*.

In 1953 the LCC published its own **Suggestions for Teachers** in comprehensive schools, which included much that the ENEF booklet had recommended, notably the need for a tutorial system for pastoral care and guidance, the development of the school as a social community in close relationship with the parents and community it served, and the opportunities that a unified secondary system would provide for new educational ideas and curricular development.

## A weaker model

In two significant matters the **Suggestions** differed from my recommendations and practice. It opposed the idea of a separately organised lower school within the comprehensive and thus hindered a salutary development: the continuity of organisation and methods upwards from the primary school, free from the downward pressure and preoccupations of 16+ examinations; the continuing role of the class teacher; a socio-disciplinary climate in which children of 11-13 thrive, and correspondingly, when they have outgrown it, a regime for the young adults of 14-18 that does not cause disaffection.

The LCC **Suggestions** also recommended no differentiation before the fourth year. While the leaving age remained at 15, this missed the chance of orientation before the year of decision. Wandsworth aimed at five years of secondary education for all. To motivate pupils towards that, we introduced a tentative bias or 'special interest' into the general course in the third year that fixed no destination but gave them a sense of purpose and direction and the knowledge that the school had something for them for which it would be worthwhile to stay.

In an article in *Forum*, vol 5 no 1, I was able to report percentages staying on in the fifth and sixth forms so conspicuously above the rest of the London comprehensives as to vindicate the obstinacy of going my own

way. This was by no means due to a favourable intake. The figures referred to the first comprehensive entry in 1956, when Wandsworth School found itself in the delimited area of Elliott School while its own area stretched away to the east to Clapham Junction, and initially produced well under 200 applicants for 420 places, the shortage being made up by 'second choices'. I remedied this administrative logic by persuading the Authority in their scheme for 1957 to allot a common area to Mayfield, Wandsworth and Elliott, that formed a fairly well-defined community on the SW boundary of London and gave parents the choice of co-education or single-sex school.

During the years of gradual expansion Wandsworth had more than doubled in size, spread its intake over a wide spectrum, and solved many problems on the way. One vital problem remained: that of the pupils we had not yet reached, with IQ below 80: the backward, retarded, or disturbed, with some non-readers, some delinquent. Our area was not short of these; but some, being unplaced, were brought in from outside it.

The large school has the advantage that it can make staff available to do justice to its minorities — if it can find staff so qualified. Much seeking, attractive allowances, and a year's secondment of volunteers for courses in Child Development enabled us to build up a remedial department of 8, who organised their own in-service training and formed a centre for remedial teachers from other schools.

For another minority who did not fall into this category but who by their fourth year had found no motivation for staying at school, we tried out methods that are described in 'Educating the Non-Scholastic' *Forum* vol 14 no 1.

## A confident start

When we finally became fully comprehensive in September 1956 the organisation that had been shaping for so long in staff discussions slipped into gear from the first morning. We had gone so far already. There was confidence that things would work. But confidence was born of more than organisation. I shall briefly describe three among developments already well-established that gave assurance that we could meet problems that lay ahead. They deal with situations that continue to challenge the schools, and one learns from experience, even other people's experience.

It was the displacement of the old Form Master by the specialist teacher that led to the emergence of the tutorial system. The Board of Education first noted it in their 1937 Annual Report and the Spens Committee recommended its wider adoption in 1939.

Appreciating the need, the totally specialist Staff of Scarborough High School put into practice in 1926 a tutorial system to which experience soon gave the following features.

## Tutor sets

The school was divided vertically into families of about 30 boys under the permanent tutelage of a master who of course added this role to his teaching duties. The Tutor Set representing the whole school in microcosm had its home room and lockers in the Tutor's class room. The home room was the base from which they went to their classes, where they assembled morning and afternoon for registration (with Board of Education permission!), and where all the usual routine business was done. Once a week they had an extended assembly and once a fortnight a last period tutorial session, which included a review of the

assessments by his teachers of each boy's diligence on a simple 3-point scale that replaced the old class marks and form orders and in the aggregate was more illuminating to the tutor, and also fairer to the boy since it put the dull and the clever on equal terms.

## Parental involvement

Home and School relations are an essential part of a tutorial system. Tutorial at-homes brought large numbers of parents to the school for each tutor to meet collectively and whenever needed to interview individually. The school was always able to provide interesting occupation for those awaiting their turn. Every meeting of parents concluded with 'open forum' at which any matter relating to the school might be raised.

Having no doubt that parental interest and support was a main factor in enabling a boy to gain full advantage from what the school provided, I welcomed the formation of a Parents' Association — rather a novelty in those days, especially in secondary schools. Parental views and attitudes are most readily developed in the group. We encouraged the Association to become an active and responsible body within a constitution that safeguarded the rights of the Authority and Governors, excluded party political and religious sectarian debate, but included all aspects of school life, and educational and social matters generally.

The tutorial system had similar effect at Wandsworth, where the Parents' Association elected an Executive Council which set up social and catering committees to serve both the parents and the school, a help fund to provide milk, meals and outfits in cases of hardship, and generously supplemented the school's provision and amenities. Of particular interest was the Exploratory Committee that met to study in co-operation with appropriate members of the staff any matter or situation that required clarification.

The Council's essential job was to plan the annual programme that brought parents and teachers together in discussion with each other and with people concerned with various aspects of education. A perennial concern was vocational guidance and careers. In this field they were active in bringing expertise and useful resources from the local community as well as from further afield.

The ENFF Booklet of 1950 had a section on Technical Education and the Link with Industry. It spoke of 'breaking down the false and mischievous dichotomy between vocational and cultural education.' The incorporation of a technical school in 1947 gave Wandsworth an organic link that was in due course substantially strengthened.

## Careers guidance

Preliminary work of a kind had been done in the Grammar School. In 1926 at Scarborough the Careers Master was

among the first to hold such an appointment. In the absence of anything much else, he linked up with the newly formed education section of the National Institute for Industrial Psychology (NIIR) and formed the nucleus of a Careers Library from its Quarterly, **The Human Factor**. After arriving at Wandsworth I appointed a Careers Master who continued in that office for over thirty years. He brought great enterprise to the job, using the resources of the Headmaster's Employment Bureau, arranging recruitment schemes with particular firms, getting help from Rotary vocational service, Chambers of Commerce, Parents, and Old Boys. In 1936 four headmasters — yes, the same — investigated the post-school careers of all pupils who had left in the previous five years and published a disturbing report on the degree of failure to have received or sought guidance. This was widely studied in educational, business, and professional circles.

A chapter on Vocational Guidance that I wrote for the **International Yearbook of Education** (1954) gives an account of the pre-war years and the establishment after the War of the Youth Employment Service.

Amalgamation with the technical school brought the Chairman and Vice-Chairman of the Brixton School of Building, representing both sides of Industry, on the Governing Body, established a firm link with Further Education, exposed all pupils to an environment enhanced by industrial expertise in Building Workshops (and later Engineering), staffed by teachers with stipulated industrial experience, who remained in touch with firms they knew and with the boys they sent to them as apprentices.

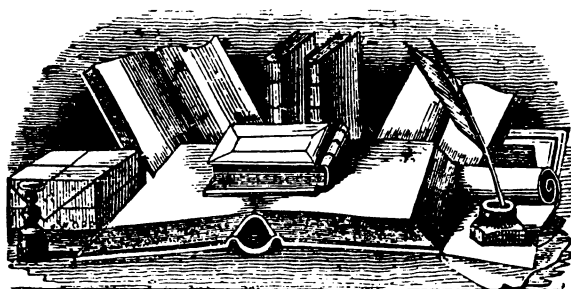
An Advisory Committee was formed representing the Building Industry and the Technical College, and then a similar Engineering Committee, who approved the technical courses and accepted the School's internal assessment in lieu of external examinations for craft apprenticeship courses in the Technical College.

## Community support

Industry was represented as well as leading community interests in School and Community, an influential group I got together in 1955 as the three great local comprehensives were on the point of opening, and which continued in somewhat modified form until the end of my time.

Parental opinion had been carried with us over the gradual expansion, so that when the anti-comprehensive movement started up in Wandsworth it fell completely flat.

A final thought. Had internal and external relationships at Wandsworth a generation ago been institutionalised in its governance, a balanced partnership of teachers, parents, community representatives, and the LEA would have been the fitting pattern.



# What and Who is the APU?

Joan Simon

Joan Simon, occasional reporter for Forum, investigates the Assessment of Performance Unit.

'Efforts are being made at every stage to inform everyone in the educational service and in schools of its activities'

So wrote the head of the Assessment of Performance Unit in the Department of Education and Science two years back in a teachers' journal.<sup>1</sup> There has been a certain flow of information and comment about the tests it is sponsoring to assess the outcome of schooling on a national scale — initially tests bearing on 'language' and mathematics commissioned from the National Foundation for Educational Research, and on science from teams at two universities. But few know what exactly the APU is, or who has been directing work which must closely affect the schools and might undermine the long established pattern of rights and responsibilities in the education service. Hence a brief from the editors of *Forum* to dig up and set out the basic facts under a title borrowed from Edward Lear

'Who or why, or which or *what*

Is the Akond of Swat?'

Nowhere could I find a straightforward account of the APU's origin, even its date of birth seems unrecorded. Some time in July 1975 is a near enough guess, perhaps, but there was a preliminary announcement a year earlier and the carefully constituted Consultative Committee did not meet until April 1976 — largely, it seems, because hard bargaining was needed to secure adequate teacher representation. As for annual reports or accounts, it is merely said that the unit is financed out of the DES grant, and the Department has always been singularly uninformative about the funds it allocates to research.

If you want the official version of operations and intentions write to the DES, Information Division, Room 1/27, Elizabeth House. You should get, in about six weeks, a series of simple little folders — apparently addressed to the general public — and some more substantial pamphlets covering specific aspects of testing and the organisation of the unit.<sup>2</sup> Any idea that the enterprise is squarely in the hands of the department of state responsible for education is deprecated in **DES Report on Education No 93** 'Assessing the performance of pupils'.

'The APU is not a group of national administrators working in isolation. Much of its work is carried out or guided by groups composed of teachers, in primary and secondary schools, local authority advisers, researchers, staff of colleges and departments of education, and members of HM Inspectorate'.

To go more directly to the point the APU comprises, at ground level, three HMIs seconded to this particular work who, with appropriate administrative backing, are located in Schools Section III of the DES; the new section concerned with curriculum and examinations, witness to considered

moves into a region with which the central authority used not to meddle.<sup>3</sup> Between them the three HMIs chair all the component groups of the APU dealing with particular aspects, from mathematics to morality. As for members of these groups, all are nominees of the DES rather than experts in particular aspects named by their peers.

Their work is directed by the APU Coordinating Group, an anodyne name for what is effectively the executive or seventeen man board of directors. No less than ten of these are HMIs or DES personnel, one is from the NFER, leaving a mere half dozen places. For these individuals were selected from a spread of institutions — teachers from a primary school, two secondary schools, a college and a university department of education, and an LEA adviser.<sup>4</sup> So much for personnel, by comparison with summary descriptions which may suggest representative or expert groups freely operating with friendly assistance from the odd HMI and servicing from the DES. What of policy, then, or the principles informing it? In June 1975 an article was published in the official journal **Trends in Education** 'monitoring pupils' progress', ostensibly a casual contribution to discussion by an individual HMI. The author, B.W. Kay, is described as a staff inspector in classics with general responsibilities for secondary education 'at present associated with the work of the APU. In fact this was the man in charge of getting things off the ground.'<sup>5</sup> Moreover, the argument set out in the Trends article turned out to be the prepared creed of the unit of which Kay shortly figured as head until 1977; he continued to chair its Coordinating Group to date.

## Six mystic measures

The main idea advanced — in an *ad hoc* way without theoretical discussion, least of all any reference to the psychology of learning — is that six 'key forms of development' can be distinguished underlying the whole curriculum. It is these that should be measured — rather than attainment in given subjects in the curriculum — if the outcome of schooling is to be adequately assessed. The 'forms of development' itemised, which supposedly add up to an overall response to schooling or indicator of efficiency in encompassing objectives, are verbal, mathematical, scientific, and physical, aesthetic, ethical. But it is far from clear what is to be measured, especially when it turns out that the related APU groups are headed 'language' (said to cover both writing and reading), mathematics, science, in true curricular manner.

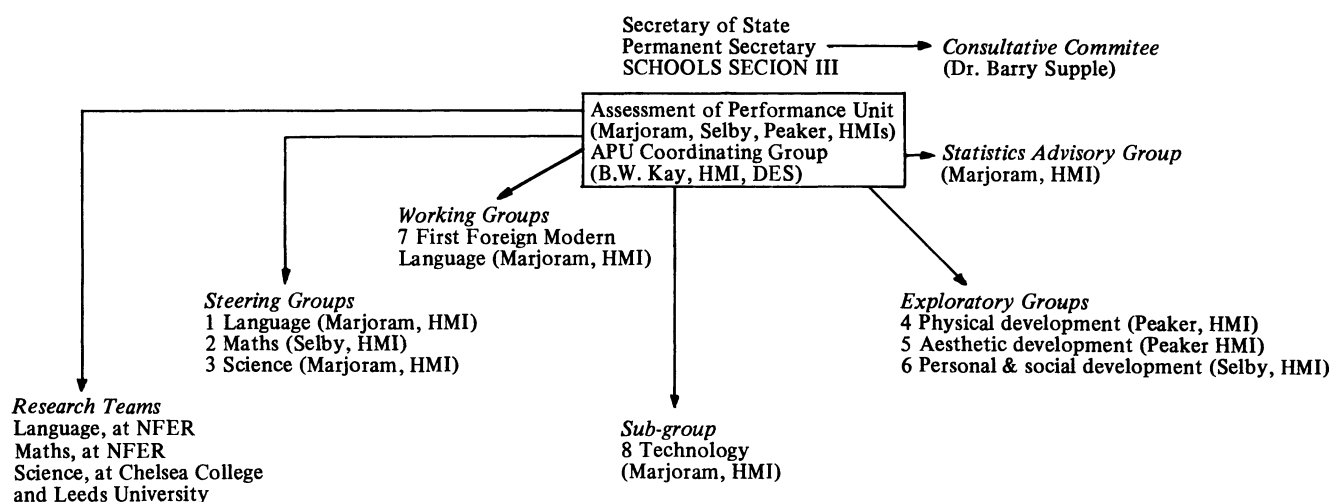
Groups established by the APU move up a scale as they realise stages in procedure, or policy. Initially called 'explora-

tory', implying a free decision as to whether it is desirable and feasible to have tests in (or of) a given area, they are transmuted, once an affirmative is forthcoming, into 'working' groups to 'map out' the territory. Finally, when the map produced finds acceptance, the DES commissions a research team to construct tests relating to the given area and the APU group acquires a 'steering' capacity. Each group numbers about eight members under the tutelage of two HMIs, selected individuals, once more, from a spread of institutions. It is notable that psychologists seem conspicuous by their absence in the inner Councils, a subject right up their street, although the odd sociologist and philosopher have apparently proved useful. Perhaps the best way of making clear the pattern described is a rough diagram of the position by May 1979.

language'. By no stretch of imagination can this be classified as a 'form of development', nor as underlying the whole curriculum. With the opening of this door, and the evaporation of the only principle supporting the plan of work, no grounds remain for refusing other applicants a hearing. What about technology and design, for instance, of key importance and hardly to be covered under the headings science or aesthetics? Hence an eighth group, or, rather, a 'technology sub-group' which consists mainly of members of other groups considering this additional claim; an indication how ingrown an enterprise of this kind can become.

Of another order is the Statistics Advisory Group which has ten members, eight of whom belong to the establishment. In this case two outside experts were invited to serve and

### DES Assessment of Performance Unit, May 1979 Components and Chairman



This shows, at the centre of affairs, the Coordinating Group chaired by Brian Kay and the trimvirate of resident HMIs — D T E Marjoram, who succeeded Kay as head of the unit in January 1977, C H Selby and G T Paker — who figure as chairmen of all the component groups. It also indicates that three of the groups set up in line with the six-point programme have advanced to the level of having satellite research teams while three remain at the exploratory stage. The exploration of ways to measure the outcome of schooling in terms of 'aesthetic' development is clearly a move into uncharted seas, but the most contentious step has been the establishment of what is called the 'personal and social development' group. According to the relevant DES leaflet this is actively exploring how to assess 'the pupil's understanding of himself, his development as a responsible person, and his moral response to his social and physical environment' — a polite pointer to proposals which extend to requiring assessment of the religious and political outlook of children.<sup>6</sup>

## The web widens

Two other groups need some explanation. It has been the official contention that all the objectives of schooling must be encompassed if assessment is to be viable; in brief that the six-point programme must be preserved inviolate. But this stance, no a very secure one to start with, has been fatally undermined by the intrusion of a seventh group, already at the 'working' stage — 'first foreign modern

both have strongly criticised the approach to test construction favoured, an aspect discussed in later articles in this issue.<sup>7</sup> Possibly it was the intention to smother such differences within the committee room; if so the plan has failed. As for the members of other APU groups, supposedly capable in due course of 'steering' the work of research teams, it seems doubtful that many are qualified to understand the highly technical questions at issue.

## Representational hoax

Finally, although the standard accounts usually put it first, there is a Consultative Committee, the only body connected with the APU which is in any way representative. It has just over thirty members, about twenty of whom are appointed by teacher and local authority associations, others being nominees of the Secretary of State. Someone had the bright idea of gathering in the editor of the *Times Educational Supplement*, another ploy which may have cramped investigative reporting of APU affairs. The non-executive role assigned to this body is summarised, aptly enough, in a DES leaflet — it 'examines the broad outlines and priorities that are being proposed for the work and brings its influence to bear on them'. No guarantee that anyone in charge of planning the programme and priorities will pay attention.<sup>8</sup>

At its first meeting, in April 1976, this relatively high-powered assembly was presented with the six-point programme for approval, learned that two groups were already

functioning and was told that the next item on the agenda was to appoint the exploratory group to investigate the ethical outcome of schooling. Not surprisingly, strong doubts found expression which have never been allayed. In short, the programme pursued under the direction of the APU Coordinating Group has never gained the approval of the Consultative Committee established to maintain essential links with the educational world. It is surprising how successfully this basic dissension has been veiled. In effect responsible representatives of professional associations have been forced into the position of questioning preconceived, if ill-digested ideas — only to find that, by the next meeting, these have been acted upon regardless. So it was that, despite cogent objections which remain unanswered, the DES went ahead to select the group to sponsor the testing of children's moral, religious, political and social attitudes, which has been at work for over two years.<sup>9</sup>

## Exit Lord Alexander

By now it will be apparent that the APU, far from being geared to tell all about itself, has been constituted on time-honoured lines; in such a way as to confine a real knowledge of what goes on to immediate participants, most especially if there is any dissension. When matters are so conducted the only option open to a serious objector is to resign, in as emphatic a manner as may be. This was the course taken in December 1977 by no less a member of the Consultative Committee than Lord Alexander, a nominee of the Secretary of State. By all accounts it was a dramatic occasion for this doyen of the LEA world abruptly left a meeting the more decisively to dissociate himself from plans to exercise central control over the curriculum and introduce tests of children's beliefs, a move foreign to the whole tradition of the education service.

No one made much of the story, however, and the crisis was damped down.<sup>10</sup> Nevertheless dissension inevitably smouldered on given continuance of the disputed group and the nature of its deliberations. These have recently been summarised in a thick duplicated document inappropriately labelled 'confidential' which ought, in all conscience, to be published before any further steps are taken.<sup>11</sup>

Barely a year after Alexander's resignation, in November 1978, one of the DES nominees on the Coordinating Group — John Eggleston, the sociologist who holds a chair of education at Keele — unashamedly urged in a public lecture that the APU

'must press ahead with testing children's personal, aesthetic, moral and physical development'

or the APU's whole plan of work would founder and it would be reduced to measuring no more than science, maths and 'literacy'.<sup>12</sup>

## Open dissent

If this looks like a put up job to muster outside support or accord an academic accolade to APU aspirations, it was soon followed by another statement, in January, which brought dissensions into the open. This, from the executive of the National Union of Teachers which has five members on the APU Consultative Committee, firmly opposes continuance of a group designed to sponsor tests of what is called personal and social development. 'Apart from the impracticalities,' it runs, the concept of apply-

ing tests on the lines proposed is 'educationally unsound and politically undesirable' and parents would certainly add their objections did they know of the project. 'The status and credibility of the APU would be enhanced by a decision not to proceed'.<sup>13</sup>

A similar view was voiced by Professor Jack Wrigley, a psychologist who knows more than most about the intricacies of assessment, both on the Consultative Committee of which he is a member and outside it. He also warned that the APU is now in danger of 'trying to measure everything', at a conference on assessment held in March at Reading University where a knowledgeable audience showed considerable scepticism about the APU activities outlined.<sup>14</sup> Small wonder, perhaps, that the Coordinating Group took the step of excluding the Consultative Committee from the annual conference held early in May to consider progress to date and 'the way ahead'; nor has any report of the proceedings appeared.<sup>15</sup> But one small ripple in its wake indicates official determination to remain head down on course — the appearance of an advertisement seeking a research assistant for the APU 'exploratory group on personal and social development'. Especially needed, it appears, is 'a comprehensive review of existing research' of relevance. It might seem inconceivable that an *ad hoc* committee should spend so long mulling over the matter without first reviewing the findings of related research, were this not the very hallmark of HMI procedure.<sup>16</sup>

## Twin birth

This outline of the position down to the close of May 1979 provides answers to the questions in my title. But another brief look should be taken at the beginnings of the APU and its terms of reference by comparison with the role now assigned to it. While official accounts say the unit was established in 1975, without explanation or greater precision, it is sometimes noted that there was a preliminary announcement in August 1974, in the White Paper **Educational Disadvantage and the Educational Needs of Immigrants**. This turns out to consist of DES comments in response to a report on education by the Select Committee on Race Relations and Immigration, for the most part defensive and directed to establishing that educational disadvantage of all kinds must be treated alike. To this end, it is said, an Educational Disadvantage Unit will be established and the terms of reference are set out in Annex A. It is in this remote corner that the APU first finds mention.

Among other tasks, the EDU is required 'to influence the allocation of resources' in the interests of 'those identified on the best currently available criteria, as suffering from educational disadvantage'; also 'to develop in association with the Assessment of Performance Unit (APU)\* other relevant criteria to improve this identification'. The asterisk directs attention to Annex B where 'for convenience' the directives for the APU are set out. Its terms of reference are (with my italics):

'To promote the development of methods of assessing and monitoring the achievement of children at school, and to seek to identify the incidence of under-achievement.

Four related tasks are enumerated: to appraise present instruments and methods of assessment, sponsor the creation of new, promote assessments in cooperation with LEAs and teachers, and, finally,

'to identify significant differences of achievement related to the circumstances in which children learn, including the incidence of under-achievement, and to make the findings

available to those concerned with resource allocation within the Department, local education authorities and schools'.<sup>17</sup>

In 1978 DES Report on Education No 93 prefaced its account of current APU work with these terms of reference and tasks without deigning to notice the discrepancy. Elsewhere they are ignored in favour of some fresh statement about the APU's role, or the reason for its foundation, related to the particular pressures of the time. Thus a DES leaflet opens:

'Everyone wants to raise educational standards. But how do we know what the education standards are? . . . The last ten years have seen changes in school organisation and curriculum. We need to be able to monitor the consequences for children's performance in school'.

That's why the DES set up the APU. This reads a bit like a response to Black Paper propaganda that there has been a general decline in standards which provided a base from which to demand continuous testing throughout school life. All right, tests there shall be but not of the kind proposed. For the proposition was tests of attainment, primarily in the basic subjects, a philistine one particularly associated with the name of Rhodes Boyson. Anyway it must have been this that was in mind when the 1975 Trends article, advocating the six point programme, warned that failure to accept it might produce

'growing pressure for the imposition of far cruder measures of educational effectiveness that could both restrict the freedom of the teacher . . . and distort the work of schools'.

## Accountability shrine

The latter points are reiterated in the 1977 Green Paper when it rejects outright the proposition that 'tests of basic literacy and numeracy' of a national character be universally applied. The alternative aim advanced — and here the APU finds mention as a natural development of the supervisory work of HMIs — is to work towards

'a coherent and soundly based means of assessment for the educational system as a whole, for schools, and for individual pupils'

in the light of a growing recognition of the need for schools 'to demonstrate their accountability to the society which they serve'.<sup>18</sup>

In sum, an initial emphasis on discerning and seeking a remedy for 'under-achievement' has given way to something more like a call for enhanced managerial control, under the convenience flag of 'accountability', with the aid of the latest in statistical devices. Even the head of the APU has cited as its main task

'to provide information which might help to determine national policy including decisions on the deployment of resources'.<sup>19</sup>

If this is really the central aim, to accumulate information facilitating the allocation of resources to match current policy — the relative lack of concern about the educational viability of tests, and of the six point programme, becomes understandable. What matter so long as results come in to the DES which appear to provide 'a far fuller and better balanced picture of educational performance

over the whole country than we have ever had before', the prospect triumphantly opened up for the public.<sup>20</sup> This may exhilarate administrators, even test constructors, but from an education point of view must seem an essentially linked aim to which to harness research funds and effort, let alone so many unpaid helpers, in the age of the micro-chip.

## New personalities

In this space reserved for any last minute postscript required it must be noted that, at the close of June, the personnel of the APU underwent a change. The head of the unit since January 1977, Tom Marjoram, was moved to the post of divisional inspector for the Metropolitan and South Midlands area — not, it was reported, the equivalent of being sent to Siberia — while the association with the APU of its original progenitor and head, Brian Kay, also ended. A fresh pattern supervened insofar as the new head of the unit, John Graham just promoted senior inspector, is responsible only for the professional aspects of the work. The administrative aspects will fall to Miss Jean Dawson (formerly dealing with adult education in HFE Branch I) who will also look after the administrative business of the Schools Council.

So much for the bones of the matter. A good deal more than a postscript is needed to consider possible developments in the 'through the looking glass' world introduced by the general election, in which Rhodes Boyson is to be found with a seat and a say in Elizabeth House. But rumour has it that while Mrs Thatcher's administration, is considering stringent 'economies' in many aspects of education, the service in which she cut her ministerial teeth, at least a million pounds has been found to ensure that the APU continues to sponsor the production of tests.



## References

- 1 D T E Marjoram, 'Assessment of Performance Unit', *Secondary Education*, vol.7 no.2, November 1977, p.9; a special issue of this journal on assessment.
- 2 The DES leaflets so far number six: **An introduction, Assessment why, what and how?, Facts and figures about monitoring, Reporting on performance** and two more on language and mathematics. The pamphlets on **Language Performance** and **Monitoring Mathematics** include lists of the APU groups involved and the research teams; a third is **Assessment of Scientific Development**, at present out of print, relating to work in hand at Chelsea College Centre for Science Education and the Department of Education at the University of Leeds. Organisational aspects are described in the occasional publication **DES Report on Education No 93**, August 1978.
- 3 According to a recent study of the DES 'the new Schools Branch III', is concerned with curriculum and examinations including liaison with the Schools Council and such related matters as educational and careers guidance and educational technology. Schools Branch 11, created in 1972, is responsible for policy on deprived and immigrant children, welfare benefits, the handicapped and special schools, while the original Schools Branch now deals solely with supply, organisation and related matters. Sir William Pile, *The Department of Education and Science* (1979), p.221.
- 4 The nominees are G H Bond, headmaster, Marshlands Primary School, Hailsham; I R Holmes, Thomas Tallis School, London; Mrs K Mitchell, headmistress, Pimlico School; WR Stirling, Bingley College of Education; D J Willmer, Warwickshire Education Department; S J Eggleston, University of Keele. Add Dr Clare Burstall, deputy director NFER, and ten internal personnel including Brian Kay, HMI (chairman), the three HMIs seconded full time to the APU, four more HMIs two of whom come from Scotland and Wales respectively, and two DES officials, A E D Chamier, B D Cullen. This list, and others given below, dates from March 1978, the last complete listing of members of all groups available.
- 5 B W Kay, 'Monitoring pupils' performance', *Trends in Education*, June 1975, pp 11-18. A DES press notice dated 15 July 1975, included the information that APU 'will operate within the DES under the direction of one of HM Inspectors, who has been responsible for the preliminary planning'.
- 6 The production devices from the leaflet **Assessment Why, What and How?**
- 7 This advisory group comprises four HMIs, two of them APU men, one from the DES, three from the NFER, plus Professor H Goldstein of London University Institute of Education and Dr D L Nuttall of the Middlesex Regional Examinations Board. Last May the latter wrote to *Education* (115.79) deploring the APU's dependence on the Rasch model. 'The testers may dazzle us with the sleight of hand, but must not be allowed to delude us into believing that they have found a magic answer to the very difficult problems of educational and psychological testing, and the measurement of standards'.
- 8 Members of this committee nominated by teacher associations number 11: 5 from from the NUT, 2 UWT, 2 Joint Four, 1 NAHT, 1 NATFHE. From the LEA side the representation is 3 Associations of County Councils, 3 Associations of Municipal Authorities, 1 Welsh Joint Education Committee; but there are two more education officers, from Cambridgeshire and Devon. Of five from universities and research institutes, one is the chairman, Barry Supple, formerly professor of economic history at Sussex University now at Nuffield College, two others are Jack Wrigley, professor of education at Reading, A Yates, director of the NFER. There remain members attributed to the CBI, TUC and National Council of Parent Teacher Associations, a director of Macmillan Publishers Ltd., the principal of a Roman Catholic college and the editor of the TES. The four assessors have included one from the Schools Council, P Dines, and P H Halsey, DES, B W Kay HMI and the chairman of the APU executive, D I Westlake, Welsh Education Office. The APU leaflet cited is **An Introduction**.
- 9 Moral education is a field in which few have been engaged and only one member of the APU committee is known to those active in it, J R Ungood-Thomas, HMI. Others listed are C H Selby, HMI (chairman); Dr C Bailey, Homerton College; Miss D Barlow, Kent Education Department; C V Surridge, Brockenhurst College; Miss R Howden, Earl Marshall (School) Campus, Sheffield; DP King, Devon Education Department; Professor R A Pring, Exeter University School of Education; K Sandhu, Sikh Education Council; Dr R Stradling, Birkbeck College; J Sutcliffe, University of Cambridge Department of Education.
- 10 'The argument is far from resolved and the ripples felt by Lord Alexander's resignation have yet to reach the bank'. noted an article in *The Teacher* (9.12.77) but the fact 'that no one has resigned from the working groups must only help the unit'. The Secretary of State attended the next meeting of the Consultative Committee and promised to pay attention to its representations in future.
- 11 One paragraph runs: 'It is the group's hope and intention that its work should secure and contribute to extensive professional and public discussion of the difficult matters with which it has been concerned'. Exploratory Group on Personal and Social Development. Report of Discussions 1976-8, p.10. It seems improper to allow this sentiment to remain veiled in secrecy.
- 12 *Education*, 10.11.78. As the lecture was delivered only to the south-east region of the British Educational Administration Society it seems likely steps were taken to ensure at least this publicity.
- 13 Comments of the National Union of Teachers on the Report of the Exploratory Group on Personal and Social Development (APU), p 5; a duplicated document.
- 14 Representatives of the APU 'did not get quite the kind of favourable response . . . they might have received in other venues on the APU roadshow', reported *Education* (23.3.79) referring to occasional planned sorties to keep the world informed. There was 'considerable scepticism. . . about what the surveys actually achieve', examples of survey questions failed to allay doubts and reports of the mental arithmetic for 11 year olds 'left acutely embarrassing silences'; much more scepticism is likely 'if the APU decides to measure personal and social development'.
- 15 The titles of key papers for the closing stages of the conference are 'The Six Lines of Development model - four years on', '“Cross-Curricularity” - what is it?', 'The Way Ahead - some questions'.
- 16 The advertisement invited applications 'to undertake a research study connected with preliminary work undertaken by the APU Exploratory Group on Personal and Social Development' including a review of existing research 'in the areas outlined' and 'critical evaluation of methods, analyses and conclusions. Other possible techniques of assessment could also be examined'. Regular consultation with members of the APU group is specified and completion by August 1980 in terms of 'a report submitted to the Group', *Education*, 18.5.79. By this means the DES correction is barely mentioned.
- 17 **Educational Disadvantage and Educational Needs of Immigrants. Observations on the Report on Education of the Select Committee on Race Relations and Immigration** (Cmnd 5720, August 1974), p.16.
- 18 See the leaflet 'APU An Introduction'; *Trends in Education*, 1975, No 2, p 18; *Education in Schools A Consultative Document*, (Cmnd 6869, July 1977), p.16.
- 19 *Secondary Education*, vol 7, no 2, November 1977, p 6
- 20 In the leaflet 'Assessment, Why, What and How?'. See the comment on DES encroachment in *Forum*, vol 20, no 1, Autumn 1977, p 18.

# The world of APU

---

## Edgar Stones

Edgar Stones, Professor of Education at Liverpool University, has taught in primary, secondary and further education and in a college of education. His main interest is in the application of psychological principles to teaching and he has recently published a book on the subject, *Psychopedagogy, psychological theory and the practice of teaching* (Methuen, 1979).

---

Clean linked go-getting CEOs with their noses to the wind and eyes to the main chance are eagerly assisting the DES and the APU in their efforts to drag us kicking and screaming into the nineteenth century. In the world of APU we make children grow by measuring them; the oftener the better. Perhaps only in England could an assessment of performance unit be established before an improvement of performance unit.

The new passion for monitoring may well go a long way towards relieving the withdrawal symptoms of eleven plus junkies but it will be no more helpful of children's learning than selection was. Much the same can be said about practically all examining as we know it.

The trouble is we all know it differently and partially. Even the most sophisticated measurer is prone to basic conceptual errors that transform the foundations of his psychometric edifice from bed-rock into sand. Less sophisticated measurers are likely to compound error by the psychometric naivety of their practices.

## Basic test offences

In examining at all levels in a vast range of courses the most elementary psychometric solecisms are perpetrated. One of the most common is the aggregating of marks from different tests. This elementary error is discussed and warned against in all primers of test theory, and almost universally ignored. The reason one should not add scores from different tests is that if the distributions of scores on the various tests differ (and they always do) the tests with large distributions affect the overall score more than those with small distributions. Thus looking to aggregate score for information about the 'clever' children will provide an answer at least as much related to the nature of the distributions of scores as the ability of the children. To get a true picture of a child's ability we would be advised to scrutinize the scores on the various tests separately.

However, a genuine searcher after truth cannot in all conscience accept even those scores. If he does he assumes that each question is of the same difficulty as every other question. This is another problem discussed in elementary texts on testing but the reaction of authors to this impasse is usually disingenuous. A synoptic paraphrase of their remarks might be: 'Yes, we accept the illegitimacy of this practice, however for our purposes in test construction we think it a defensible procedure so we will proceed as though marks on different questions are in fact additive'. And with a bound Jack is free. If to the problem of the dubious legitimacy of adding scores we add the complicating factor that in most exams candidates have a choice

of questions, it becomes clear that it is virtually impossible to categorize children accurately as the present fashion seems to demand.

But this is not all. I have so far only touched on some of what seem to me key technical shortcomings in current exam procedures. There is an additional overriding problem; that is the question of validity. Do exams in fact do what they are supposed to do? That is, do the questions we ask actually identify excellence in any given field? In all but the simplest kind of learning, for example learning to recite the alphabet, we cannot fully test the ability we are interested in, all that we can do is to take the candidate's answer as a legitimate surrogate sample of his ability. In most cases we have absolutely no data one way or the other as to whether exam results tell us anything at all about levels of competence in a particular ability. Very often we haven't even much of an idea what competence consists of. Is there, for example a recognized consensus on what a 'good' doctor, dentist, teacher, lawyer, psychologist, university professor, politician, or novelist is? Or is it not rather, the fact that criteria are derived circularly and identified as the ability to succeed in exams.

There is one further point. Conventional approaches to examining are nothing more than elaborate attempts at ranking candidates. Even if test procedures were free from errors alluded to above, all they would or could tell us is that A scored more than B who scored more than C and so on. Unless there is some objective external criterion to which the test can be related we have no means of knowing whether the standard of achievement in the test is being upheld or not. The standard of binocular vision of the one eyed king in the country of the blind would not impress me but it would impress his subjects.

## What do tests tell?

However it would be a mistake to reject tests and exams altogether on the grounds that they are used as grading devices to signal to people that they are superior or inferior to others. Forms of evaluation may be available and conceivable that will help people to learn. In rejecting grading we need not reject guiding.

The graded word reading test exemplifies the limitations of the grading function of examinations. The most commonly used of these tests was devised many years ago using statistical procedures to ascertain 'standards'. A word list graded according to difficulty as perceived by the test constructor is tried out on a sample population and norms derived from the try out. Levels of performance of the try out population are then adopted as the 'reading stan-

dards'. However there is no means of knowing whether these standards are valid tests of reading ability since the only validity they have is that derived circularly as the ability to do well on the reading test which itself was developed by seeing what children were currently capable of. Given this situation it is extremely difficult to discern any reliable grounds for anxieties about declining standards of reading. The grounds are even less firm if we consider the volatility of language usage. Tests standardized in the way described cannot but become dated very quickly as words change in familiarity, as the Bullock report points out. Thus if two or three words on the test become fashionable, possibly because they appear in advertisements or on television, reading standards are likely to 'improve' as compared with earlier years, whereas the opposite will happen if words on the test drop out of use. In essence while we are far from sure what reading ability is, most existing tests tell us little other than that some children do better and some do worse when they take the test.

The situation in other fields of testing and examining is even less impressive. Public examining boards employing cohorts of teachers to mark examination scripts every summer for the most part perpetrate all the psychometric solecisms referred to above. Some take steps to ensure marker comparability but most participants in the yearly ritual are far too busy coping with the flood of scripts to contemplate the statistical, or even educational, validity of what they are doing. In fact most of the operators of the system are pristinely innocent of any psychometric or statistical facts of life. To talk about maintaining standards in such a system is patent nonsense. Anyone truly concerned to maintain standards of competence in any aspect of schooling should be very sceptical about current procedure of examining.

## Helpful tests

On the other hand much is to be gained by considering the possibilities of alternative methods of examining or testing. We should consider the type of test that avoids the errors referred to above and which, instead of defining excellence as the ability to do better than others, see it as the ability to accomplish a task to a predetermined high level of competence.

Tests of this type comprise questions selected not because they are likely to discriminate among candidates but because there is general agreement that, in so far as is possible given the crude state of the art, they tap ability in the subject under scrutiny. (To some extent this happens with conventional tests except that the prior criterion almost invariably is the discriminating power of the test question.) In this type of test we are not just interested in ascertaining that A does better than B. This kind of result tells us very little. It tells the teacher hardly anything about the effects of his teaching since A may well have learned very little. Knowing that A has a reading age of seven when his chronological age is nine does not help a teacher to help the child improve his reading. A test based on an analysis of the constituent skills of reading is quite different. It could reveal to the teacher the areas of difficulty the child is experiencing so that both could take remedial action.

Tests of this type are extremely useful because they tell us something about a child's capabilities of a quasi-objective nature which helps with his further development. If the APU develops this kind of test I will be the first to applaud.

## Will APU tests help?

From the way things are shaping however, it seems that they are not. They are talking in terms of traditional standardised tests even though they may use the Rasch approach discussed in the next article. No matter how high pupils' achievement may be in tests of this nature about half will be considered unsatisfactory, perhaps 'under achieving' since their scores will be 'below average'. Much the same can be said about exams at 16+, 18+ 11+ or X+. On the other hand with tests that assess the achievement or non achievement of pre-agreed criterion standards the idea of 'average' is unimportant since each candidate's performance is compared with the criterion and not with other candidates' scores.

Tests that support to sort people out in a competitive struggle presumably fulfil needs expressed by an important and influential group in contemporary British society. The irony is that the rank order they produce is almost certainly 'inaccurate'. I fear, however, the irony would be lost on many, for whom the struggle avails more than the outcome.

Oscar Buros, who spent most of his professional life scrutinizing and reporting on tests of all kinds, left a message for all educators that chimes closely with what I have said.<sup>1</sup> The stress on tests that sort and categorize is deleterious to education. The current recrudescence of such testing, it seems to me, is a cancerous growth that will not and cannot 'maintain standards'; Only when we get back to pedagogic basics and bring together teaching and testing in a mutually supportive relationship can there be any chance of systematic 'monitoring' or maintaining of 'standards'.

1 Buros, O K (1977), 'Fifty years in testing: some reminiscences, criticisms and suggestions'. *Educational Researcher*, July-August, p 9-18.

## Progressive Education

**Progressive Education** is the name of a new organisation the aims of which are to bring together teachers who share a 'progressive' point of view and to provide a counter-balance to the negative trends in education which are now so much in fashion. It is organised on the basis of a number of local groups throughout the country and a Central Co-ordinating Committee. The job of the central committee is to facilitate the working of the local groups, to organise conferences and workshops, to publish a twice yearly newsletter and to provide a directory of information.

If you would like to know more about us or to be put in contact with your nearest local group, please ring or send sae to:- Lesley King, 4 Central Street, Countesthorpe, Leics LE8 3QJ. (Tel. Leicester 774577.)

# The Mystification of Assessment

---

## Harvey Goldstein

Since 1977 Harvey Goldstein has been Professor of Statistical Methods at London University Institute of Education. Prior to that he worked at the National Children's Bureau and was principally responsible for the statistical analysis of the British National Child Development Study. Here he exposes the so-called 'objectivity' of the Rasch model now favoured at the NFER in its work for the APU.

---

A few years ago, educational assessment finally started to emerge from the dominance of IQ Testing, and began to consider the need for carefully thought out *educational* as opposed to *psychological* motivations in devising assessment procedures. The subsequent development of notions such as criterion-referenced testing and process evaluation has demonstrated that a variety of tools can be devised, although many of these have been seen as providing qualitative rather than quantitative descriptions. For those who are interested in making comparisons among individuals and groups, however, some form of quantification of assessment seems necessary, and one of the principal tasks for educational statisticians, as I see it, is to provide acceptable quantitative tools for educational measurement. Nevertheless, since quantification without educational content is insufficient, and since it is often all too easy to develop the former to a level which is technically sophisticated while lacking the latter element, any new statistical technique needs to be evaluated carefully in terms of its basic premises and assumptions. The purpose of this article is to examine one such technique which has become known as 'objective measurement', on behalf of which some far reaching claims have been made.

A detailed critique of 'objective measurement' and the related 'item banking' methods in the context of national and local testing programmes has been made elsewhere (Goldstein and Blinkhorn, 1977). Here, I shall explore the underlying philosophy of this approach and try to see what implications it might have in general for educational assessment. Before doing this, however, it will be useful to make a few brief comments on certain general critiques of educational measurement which seem to reflect a rather widespread misunderstanding of this term.

In a recent article, McIntyre and Brown (1978), while making some apt comments about the use of psychometric techniques in measuring educational attainment, appear to equate attainment measurement itself with psychometrics, while they elsewhere refer to the 'assessment' of attainment as if this did not involve any 'measurement'. Of course, anyone is at liberty to suggest their own definition of a term, but it would be extremely unfortunate if a rejection of the *psychometric* approach to measurement were to lead automatically to the rejection of *any* measurement or quantification in education. Thus even the act of deciding whether or not a child has achieved a state objective can be viewed as a simple measurement process, and as soon as one is confronted with a number of such simple judgements, some logical ordering or summarisation of them usually becomes necessary.

The relevant questions to ask are those concerned with the

levels and types of measurement or quantifications which should be used to describe the defined attainments. For example, in criterion referenced testing we may wish to deal with a simple yes/no categorisation, but this then leads on naturally to the calculation of percentages of individuals responding 'yes' and to comparisons of such percentages across groups of children. None of this activity is necessarily inimical to the objectives embodied in the test used. Yet it does involve measurement and quantification, and hence the possibility of statistical analysis and summary. The statistical methodology itself has no inevitable connection with particular educational theories, even though its use has often been to buttress such theories. Thus Stenhouse's (1978) criticism of the 'psycho-statistical' approach seems to stem from his identification of particular techniques with certain approaches to educational measurement, and does not really stand up as a fundamental critique of the use of statistical methods as such.

## Outstanding problems

To say that statistical methodology has a role in educational assessment, however, is not to deny that there remain considerable problems. Two of the principal difficulties which have always faced constructors of traditional educational tests, for example, have been those associated with the need to provide a balance between test items appropriate to children exposed to different curricula etc., and also with the problem of test items becoming outdated with the passage of time. This latter difficulty, in particular, has thwarted attempts to make judgements about trends in educational attainment, since any apparent change might simply be reflecting the changing difficulty of the test rather than the achievement of individuals. Thus for example, any concern about so called 'falling standards' will have to face this issue, and it is in dealing with this particular issue that claims on behalf of 'objective measurement' have been advanced quite strongly.

Although nearly all the important developments in the methodology of objective measurement have come from Scandinavia and the United States, the British work in this area will be used as an illustration since its effects are more immediately apparent in this country. While several British educational researchers have been involved in objective measurement the most important centre of work lies inside the National Foundation for Educational Research (NFER) and most of the ideas behind the work of this group of NFER researchers are spelled out in the book by Alan

Willmott and Diana Fowles, *The Objective Interpretation of Test Performance* (NFER 1974).

The following quotations from Willmott and Fowles give a fair summary of their underlying approach to objective measurement.

‘An ‘objective’’ measurement is one of the type which is so familiar in the physical sciences.’

and to establish ‘objectivity in a test,

‘. . . requires first that the characteristics of the items in a test must somehow be made independent of the distribution of attainment in the group who are given the test, and, secondly, that the test should give estimates of attainment which are independent of the particular set of items which comprise the test.’

This seems to express a deeply felt need to make education as much like physics as possible, and even embodies the rather curious notion (in the light of the Special Theory of Relativity) that the properties of a measuring instrument are invariant, and exist independently of the circumstances in which it is used. Even so, the authors are prepared to concede that such an ideal situation may not always exist and that some of the items in their tests may not conform to their ideal. Writing about their ‘model’ (which I shall return to later) for an objective test they say:

‘The criterion is that items should fit the model, and not that the model should fit the items.’

This is an extremely radical proposal. What it says in effect is that, given the particular assumptions embodied in the mathematical formula which relates a test item score to an individual’s ‘underlying ability’, any test item which does not conform to the general pattern is simply discarded. This can only mean that educational reality is subservient and can be deformed in order to satisfy the ‘model’ rather than that the ‘model’ is revised in order to better describe any educational reality. In fact, Willmott and Fowles make no serious attempt to consider how their mathematical model might be justified by any educational model, and after they have discarded the ‘non-fitting’ items in their tests they pay scant attention to whether what is left actually measures anything sensible.

## Who are the misfits?

They even discuss what to do with those individuals who do not happen to fit their model, and talk of;

‘Some form of diagnosis on the part of those who knew the candidates well.’

In other words, the ‘norm’ consists of conforming to the ‘model’ and those who do not are considered abnormal and needing diagnosis. Thus, a value judgement seems to be implied and because the model tends to ‘fit’ individuals who have broadly similar response patterns to test items, it would not be too surprising if the ‘abnormal’ individuals turned out to belong to cultural minorities or to be those following a novel curriculum. There is no discussion of the dangers of such a possibility in Willmott and Fowles, nor, so far as I can tell, by any other proponents of objective measurement.

The ‘model’ upon which the above claims are based is known as the Rasch Model, named after the Danish mathematician George Rasch. It is quite unnecessary to go into technical detail in order to describe the more important assumptions upon which this mathematical model rests (these can be found in Goldstein and Blinkhorn, 1977). They can simply be stated as follows. First of all it assumes that there is a single ‘trait’ or ‘factor’ which determines the chance that an individual responds correctly to an item in

a test. For a mathematics test, for example, it would be assumed that something called ‘mathematical ability’ existed and could be described for an individual in terms of a single number. Every item in the test is likewise supposed to have a single value known as its ‘difficulty’. Secondly, the model assumes that the difficulty order of the items in a test remains the same *for all individuals*, whatever their backgrounds, their exposures to different curricula etc. Stated thus, the limitations of the Rasch model are fairly obvious, so that any use of it ought, at the very least, to be tentative and exploratory.

## APU adopts Rasch

Nevertheless, advocates of objective measurement are not simply spending their time playing with their models inside research foundations, where it might be regarded as a possibly interesting and fairly harmless academic pursuit. Unfortunately, this is far from being the case. The NFER, for instance, is carrying out much of the monitoring work for the DES Assessment of Performance Unit (APU), and some of their more senior researchers have been advocating the use of the Rasch model in the design and analysis of APU tests.

One particular proposal envisages a ‘bank’ containing a very large number of items which will be ‘calibrated’ against each other, and selections made according to a user’s specifications. Resulting tests, it is claimed, will be suitable for use by testers without further modification, and with a ready-made calibration available so that results can be scored on a common ‘objective’ scale. In particular, the results from tests designed for different curricula can be compared with each other, so that all individuals can be ranked on a single scale (excluding presumably those who do not fit!). Not only is such a possibility actually unattainable, it is also highly questionable as to whether such a goal is even desirable.

Nevertheless, the possibility of such an absolute measurement scale has a certain attraction and it is unlikely to be abandoned easily, despite any lack of educational relevance. A further strong reason for a reluctance to abandon it arises from its claim that it provides a method of making comparisons so long as all the items used are selected from the same item bank with ‘new’ items being calibrated against ‘old’ ones. This, however, is also impossible, even within the assumptions of the Rasch model, as the following simple argument shows.

## Illogical claims refuted

If we suppose that each of the items in the bank has a prescribed difficulty value, then it is strictly meaningless within the context of a Rasch model to speak of one item as being *more applicable* to one point in time rather than to another. The only meaning which can be attached to such a statement must be in terms of difficulty values. For example, suppose there are two items, one of which is more applicable in 1975 than 1980 and the other of which is more applicable in 1980 than in 1975. Then the two items will have different relative difficulties in these two years, and indeed their relative difficulty might become reversed between 1975 and 1980. Hence, by definition, they cannot belong to a *single* common Rasch scale extending over this five-year period. Nor will it be possible to ‘calibrate’ their difficulties via other items whose difficulties, for the sake of argument, are assumed to remain constant. Thus an item

bank based on the Rasch model and designed so that outdated items can be replaced is a self contradictory concept. A similar logic applies to claims that an item bank can be constructed to suit different curricula. Thus, despite its claims the methodology of objective measurement contributes nothing to the resolution of the difficulties facing test constructors, which were mentioned in the introduction.

## Disarming simplicity

I have argued that the objective measurement movement in education is misguided and offers over-simple solutions to complex problems. It is, however, the very simplicity of these solutions which is extremely seductive, and this is compounded by the jargon which surrounds the methodology and which has about it an air of desirability and promise. The one thing which is not revealed clearly by those who advocate the methodology, however, is just what the mathematics actually implies in educational terms. Certainly one searches in vain for such statements in Willmott and Fowles. It is in this way that the advocacy of objective measurement tends to lead to mystification. It takes the discussion of the curriculum and its evaluation out of the main educational forum, and essentially hands it over to technicians who can manipulate the mathematical equations. In the final analysis, even more than the use of objective measurement itself, it is this mystification which seems to pose the most serious threat to the enterprise of education evaluational and assessment.

### Acknowledgements

I would like to thank Tessa Backstone, Steve Blinkhorn and Bob Wood for their helpful comments on an early draft of this article.

### References

- |  |  |
|--|--|
| 1 Goldstein, H. & Blinkhorn, S. (1977) | Monitoring Educational Standards – An Inappropriate Model; <i>Bull. Br. Psychol. Soc.</i> 30, 309-311      |
| 2 McIntyre, D. & Brown S. (1978)       | The Conceptualisation of Attainment; <i>Br. Educ. Res.J.</i> 4, 41-50                                      |
| 3 Stenhouse, L. (1978)                 | Case study and Case records: towards a contemporary history of education; <i>Br. Educ. Res.J.</i> 4, 21-40 |
| 4 Willmott, A.S. & Fowles, D.F. (1974) | The Objective Interpretation of Test Performance (NFER)  |

## A Subscription to Forum

To: Forum,  
11 Beacon Street, Lichfield.

\* Please send three issues of **Forum**, starting with  
Vol \_\_\_\_\_ No \_\_\_\_\_ to the address below.  
I enclose cheque/P.O. for £2.50.

\* Please send a Banker's Order Form to the address below.

NAME \_\_\_\_\_

ADDRESS \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\* delete as appropriate

# Accountability

## – a contagious disease?

### Joan Shapiro

Dr Joan P Shapiro is a Lecturer and Supervisor of Teacher Education at the University of Pennsylvania's Graduate School of Education in the USA. She wrote this article while on a year's leave of absence (1978/9) as an Honorary Research Associate in the Curriculum Studies Department at the University of London Institute of Education.

This article is an attempt to provide British educators with a synthesis of some major developments in the Accountability movement in education, focusing primarily on the US experience. It has been written because there appears to be a need to understand a trend, which is now spreading in the UK. Since the Accountability movement in the US is advanced by five to ten years on its British counterpart, it is hoped that an understanding of the American experience may provide some insights into the present UK situation.

In both the US and the UK, the term Accountability has been interpreted in diverse ways. The varied definitions range all the way from narrow monetary concerns to broad political connotations. However, irrespective of the definition utilized, I believe that Accountability is an understandable outcome of an unstable social and financial decade. During the 1970s, inherent mistrust of most institutions have led politicians, ratepayers, and the media to assume that they have the right to hold the school accountable not only for its spending but also for certain aspects of the educational process.

The punitive tone and the negative forms associated with Accountability in the US have led me to liken this trend to a disease. If viewed as such, then the rapidity of the spread of this movement suggests the Accountability is contagious in nature. Furthermore, in light of the negative attitude exhibited by the general public towards education, it is clear that the ability to ameliorate this condition is not good. Unless controlled, it is my contention that the Accountability movement will have a stultifying effect on education, inhibiting innovation and progress in this field for many years to come.

It may be argued that an analogy which compares Accountability to an infectious disease is a gross exaggeration. It may also be debated that the US is so different from Britain that reported abuses of Accountability in America could never happen in the UK. However, I would suggest that to ignore the causes, signs and symptoms of Accountability in America would be very foolhardy. Indeed, even if only a few similarities between the two countries can be identified, then I believe that an analysis of the American Accountability movement merits some serious consideration.

Thus, in this paper, I shall discuss causes and effects of Accountability in the US; and I shall draw the reader's attention to similar developments already occurring in the UK. Finally, I will propose a treatment plan to help monitor and limit the growth of this movement.

In the US, disillusionment with state education was a noticeable aetiological factor associated with Accountability. One case for this disenchantment was suggested by Bowles and Gintis (1976). They argued that Accountability was an outcome of the public's disdain for the 'American Dream'. In the 1950s attainment of this dream appeared to be a possibility for those who obtained a university degree. By the mid-1960s, one-half of the appropriate age group had entered college. But instead of achieving their dream, these young people discovered that a restricted employment market awaited them. In the 1970s, disappointed ratepayers demonstrated their own and their children's unhappiness with the schools by failing to support and to fund state education.

Disillusionment with education has also been attributed to the public's increasing contempt for the promises of the social reformers and progressive educators in the 1960s. The societal instability of the sixties, embodied in Civil Rights marches and campus protests, seemed to create in many Americans a desire to search for, and return to, those values which they perceived to have stood the test of time. Groping for security in a very insecure society, the public, in the 1970s, rejected change and were conservatively harkening 'Back-to-Basics'. In the name of a basic education – frequently defined as the 3Rs – the public and politicians felt justified in recommending drastic reductions in school spending. Thus disillusionment with state education led to both economic restrictions for the schools and to a narrower 'basic' curriculum for students.

## Falling rolls

Disenchantment with state education was not the sole reason for the growth of Accountability in America. Another underlying issue of importance was the dramatic decline in the birthrate. Women, who had been averaging 3.50 children in the 1950s, were averaging 1.76 children by the late 1970s (Elam, 1978). With a decrease in the number of children requiring schooling, it might be expected that declining enrolments could have had a very positive effect on education. This decline might have led to an awareness by the public that here was a 'heaven-sent opportunity to improve conditions!' (Pedley 1979). Unfortunately, this view was not expressed by many people. Instead, the majority of US ratepayers regarded the decline in the number of children as a chance to cut down on their property taxes. This reaction led to the closure of many schools and the dismissal of teachers throughout the country.

The decisions related to declining enrolments were only one sign of the growing demands of ratepayers to decrease their property taxes. In 1978, the culmination of the ratepayer's rage was seen in the passage of Proposition 13. In one sweeping decision, California voters rejected local school district support of education.

## Proposition 13

To understand the impact of Proposition 13, it is necessary to have some knowledge of the funding of state schools in the US. In the American Constitution, delegation of this function was assigned to the state government. However, over the years, local government sought and assumed this role in order to achieve autonomy in decision-making for its schools. Passage of Proposition 13 removed this autonomy and returned power over the local school districts to the state.

A factor in Accountability, not related to fiscal constraints and social upheaval, is the reliance placed on assessment techniques. Educational testing in the US was developed both to monitor individual attainment and to estimate national levels of achievement in state schools. It was also associated with teacher training and teacher evaluation. The assessment movement relied heavily on the Objectives Model in education. This model emphasized prespecified objectives, noticeable changes in pupils' behaviour, and measurable results. Conceptually, such a model stressed the product or outcome of learning rather than the process of education.

Since 1975, to determine individual pupil attainment, there was a trend towards assessing minimum competency learning. In thirty-three states, specific minimum levels of knowledge, measured through product evaluation, must be achieved by each elementary and secondary student. Although the requirements varied from state to state, generally minimum competency meant that a child would be tested at different times in his/her development. If, at any stage, minimum results were not achieved, the child would be either held back a year or receive remedial help.

It seems clear that all children should achieve certain minimum levels of learning; however, the punitive aspects of this type of testing have often been ignored. The deleterious emotional effects of holding children back from their peer group, and the possible consequences of dismissal action for teachers, whose pupils have not all met minimum standards, are issues worthy of consideration.

## 'Blanket' testing

In addition to testing pupils for minimum attainment, the US was also engaged in determining national assessment levels. The infectious nature of this kind of assessment was noticeable because not only were national agencies performing educational testing, but individual states also initiated their own programmes. Thus far, it has been the states which have caused the major problems. For example, Michigan ordered 'blanket' testing for its pupils and then attempted to utilize the test results as a vehicle for dismissals and promotions of teachers (Burstall and Kay, 1978). In this way, as a form of 'payment by results', some of the worst fears of teachers almost became a reality. Only the action of an irate teacher's union prohibited that state from achieving this end.

## Effects on teacher training

A further outgrowth of educational assessment was seen in its use in the training of teachers. Performance-based teacher education has been required in seventeen states. In this type of training, it was believed that teaching could be analyzed according to types of teaching activities – ie explaining, guiding, disciplining. As such, performance functions were stated as clear objectives so that they could be subsequently measured. One of performance-based teacher education's harshest critics (Nash 1970) stated that in the name of Accountability the US was 'in danger of reducing the entire teacher training programme to such trivializing exercises as the preparation of general and specific instructional objectives' (p 241).

Another system of the times, closely associated with assessment, was the emphasis on teacher evaluation. From the early 1960s, state legislators had been passing teacher evaluation laws. Now, in over twenty-three states, these laws exist. Although there was some flexibility in the

guidelines of each state, the emphasis on product evaluation and the punitive tone of the laws were evident. Thus, teacher evaluation became an area of controversy and teachers' unions have been actively engaged in combating this type of restrictive legislation.

## Some comparisons

Although profound differences exist between the educational systems of the US and the UK, as far as the Accountability movement is concerned, parallels can be drawn with some caution.

Firstly, with respect to disillusionment with state education, which British teacher has not become aware of the comments of parents, politicians, and the press related to the perceived decline in the standards of learning of children? This decline has most frequently been attributed to comprehensive schools and to progressive education. The same critics have also expressed their desire for increased emphasis on skills and on a 'basic' education. This punitive tone and conservatism towards the schools is strikingly similar to that noted in the US.

Another similarity between the British and American scene is a dwindling school population. In the UK, as a solution to decreased enrolments, there has appeared to be some reliance on school closures. Over five hundred rural schools have already closed (Rogers, 1979), and the urban areas are just beginning to be affected. In the wake of closures, dismissals of teachers could follow.

Parelleling the US, economic problems in the UK are very real. Inflation and strikes create an atmosphere which put the public in no mood to provide educational 'frills'. In such a climate, British educators should expect to be held more accountable for the spending of public funds.

## Warning of APU

Assessment is a particularly noticeable similarity between the two countries. In England, the APU and local school authorities have been actively devising tests for pupils. What uses will be made of the tests data have yet to be discovered. However, if the US experience is used as an example, then I would predict that inaccuracies in data analysis will flourish and that media distortions in interpretations will abound.

Clearly, I believe that the Accountability movement in the US has been detrimental to that country's educational

development. Although not as advanced as its American counterpart, a similar condition is infecting the schools in the UK.

The aetiological complexity of the Accountability movement suggests that no simple treatment plan can be prescribed. Therefore, each country must devise its own appropriate ways of dealing with problems produced by the trend. However, the American experience has indicated that consumerism and politics in education is a reality and that an attitude of detachment and passivity by professional educators is unacceptable.

## The need for alternatives

What is urgently required in the UK is concerted educational leadership, from both the schools and universities, to counteract the movement. Educators, at all levels, need to criticize, monitor and provide alternatives to Accountability. It should be British educators themselves who evaluate the schools; report the findings to the general public; perform the necessary research studies to meet parental criticisms; assist and 'up-grade' the professional staff; and dismiss, if all help fails, incompetent colleagues. Only through rapid and united action will British teachers, educational administrators, and faculty in schools of education restore confidence in the state educational system. In so doing, they may be able to limit the effects of Accountability before education for a future generation is jeopardized.

### References

Bowles, S and Gintis H (1976)  
Burstall, C and Kay, B (1978)

Elam, S (1978)

Nash, R J (1970)

Pedley, R (1979)

Rogers, R (1979)

**Schooling in Capitalist America**, Routledge & Kegan Paul.

**Assessment: The American Experience**, Department of Education & Science (mimeo).

'Ways of Dealing with Enrolment Decline: A Telephone Conversation' in *Phi Delta Kappan*, Vol 60, No 1, pp 20-5.  
'Commitment to Competency: The New Fetishism in Teacher Education' in *Phi Delta Kappan*, vol 52, No 4, pp 240-3.

'Towards the Comprehensive University' in *Forum*, Vol 21, No 2 p 72.

**Schools under Threat: A Handbook for Closures**, Advisory Centre for Education.

### References for 'A Constructive Response to the APU' p.20

Barker-Lunn, J (1970) **Streaming in the Primary School**, NFER.

Bennett, N (1976) **Teaching Styles and Pupil Progress**, Open Books.

Burstall, C and Kay, B (1978) **Assessment — The American Experience**. Assessment and Performance Unit, Monograph, Department of Education and Science.

Choppin, B (1978) **Item Banking and the Monitoring of Achievement**, NFER

Eggleston, J F, Galton, M J and Jones, M E (1976) **Processes and Products of Science Teaching**, Schools Council Research Studies, Macmillan Education.

Galton M J, Simon B, and Croll, P (1980) **Inside the Primary Classroom**, Routledge & Kegan Paul, (in press).

Harlen, W, Darwin A, **Match and Mismatch: Raising**

and Murphy, M (1977)

Jasman, A (1979)

SCRE (1977)

Sumner, R and Robertson, T S (1977a)

Sumner R and Bradley, K (1977b)  
Willmott, AS and Fowles, D E (1974)

**Questions, Leader's Guide**, Oliver & Boyd, Edinburgh.

'Developing Teacher-Based Assessment for Primary Schools' in *British Journal of Educational Research*, Vol 5, No 1, pp 101-110.

**Pupil in Profile: Making the Most of Teachers' Knowledge of Pupils**, Scottish Council for Research in Education.

**Criterion-Referenced Measurement and Criterion-Referenced Tests**, NFER.

**Assessment for Transition: a study of new procedures**, NFER.  
**The Objective Interpretation of Test Performance: The Rasch Model Applied**, NFER

# Discussion

## Assessment and Pupil Profiles

A report of a conference organised by PRISE (Promotion for Reform in Secondary Education) in March 1979, to consider alternative ways of assessment other than the public examination system.

The conference began by considering a previously distributed paper entitled, 'The Myth of Comparability', which was prepared for PRISE by Desmond Nuttall, Professor of Education at the Open University. The work, based largely upon studies by himself and Alan Wilmott of NFER, argued that there is no way of achieving comparability, between boards in a given subject, between subjects within a given board, between years in a given subject in a given board, other than relying on trained, but fallible human judgement. After saying that comparability was not content comparability nor skills comparability and that therefore it was identicality of performance, the paper went on to consider whether comparability mattered and suggested that that issue paled into insignificance beside the inevitable inconsistency of the assessment process itself. In the final section it was suggested that more should be done to make it more widely known that comparability is only rough and ready. However, the consequences of this might be pressure for standardising syllabuses or prescribing the criteria to be tested in any examination. There were already pressures for this from SCUE at A level (and at N & F) and the Waddell Report moves in this direction for a common system of examining at 16+.

So a dilemma faces us. Should the genuine, but somewhat misguided, according to Nuttall, desire for comparability be pandered to by setting up elaborate moderation, validation and accreditation procedures and give labels, either in words or grades that are designed to have some common currency, or should each school be allowed to develop its own reporting procedures? This latter invites employers, colleges and universities to take much more initiative in devising their own selection procedures: in some ways this could be constructive, but it does, says Nuttall, invite 'league-tabling' of schools, nepotism and all those bad effects that public examinations were originally devised to overcome. It also implies a boom in, often inappropriate, psychological and educational testing by employers, with possible unfortunate backwash effects into the schools, let alone the creation of problems for those on the milkround of job seeking. So Nuttall urges

care to avoid leaping from the frying pan into the fire.

Questions to Desmond Nuttall and Robin Smith (Secretary of the Oxford Examinations Board) and discussion among conference participants raised such issues as: would removal of the public examinations at 16+ free the secondary curriculum as removal of the 11+ examination had liberated the primary school curriculum? Were CNA validation procedures applicable to schools? Would local industry try to impose syllabuses? If the exam boards spent so much time, money and effort in trying to obtain comparability could the schools really do things better? Do not school leavers get their jobs before the results of external examinations are known so is it not the school report or profile that is more important?

The conference then heard about three pupil profile schemes that have recently been developed. The first, by the Scottish Council for Research in Education and the Working Party on School Assessments of the Headteacher's Association of Scotland, was reported by Patricia Broadfoot, lately a Research Officer in SCRE and now at Westhill College, Birmingham.

It should be remembered that Scotland has no CSE exams so a considerable percentage of the school population is not catered for by an externally accredited examination, although as the **Dunning Report** made clear, a far higher percentage is presented for O grade examinations that was originally intended. The SCRE Profile Assessment was developed to meet the needs of all pupils for self-knowledge, for curricular and vocational guidance and for a relevant and useful leaving report. It is a diagnostic rather than a judgemental assessment based on the principles of justice, relevance and practicability.

The profiles, produced perhaps twice a year, provide not only a comprehensive and cumulative basis for within-school guidance but also the necessary information for a leaving report for each pupil, covering basic skills, subject achievements and personal qualities. The final leaving report aims to be brief, accurate, positive and useful to potential users and pupils. It is summative and does not include all the subject specific categories useful for in-school purposes. The basic skills include: listening, speaking, reading, writing, visual understanding and expression, use of number, physical co-ordination and manual dexterity; each is graded on a four point scale, each point being criterion referenced. Not all skills are assessed by all teachers. Initially there were eight personal qualities to be assessed but these have been reduced to two – Enterprise and Perseverance (or Inspiration and Perspiration!). The system is an open one that the pupils know about. It establishes assessment as communication and as a learning process. It recognises the individual and his attributes, the emphasis is on description and not comparisons between pupils.

Bob Mahy, (Deputy Head, Sutton Centre, Sutton-in-Ashfield, Notts) spoke about the scheme developed and used at the Sutton Centre. Pupil groups each have a tutor who stays with them as long as the tutor and the group are in the school. Home visits and relations with parents are highly rated. The ethos of the staff-pupil relationship is based on people being equally important. All tutor groups are of mixed abilities and the em-

phasis at the Centre is on individual work. Nowhere in the Centre does any grading or labelling occur until the external examinations are taken. Everything within the school is individualised and personal. The external exams were, until recently, Mode 3 CSE with total continuous assessment.

It was against this background that the Sutton Centre Profile was developed. What was wanted was something more than the usual report, something which supplemented the CSE results. The profile was conceived as a yardstick by which the pupil could measure himself, and a means by which others could measure him. The profile seeks to achieve four ends, to be:

*a three way dialogue between staff, pupils and parents*

*a record of a pupil's achievements, an information document useful to employers about the courses followed at the Centre.*

Information is kept in a file under the following headings:

*introduction, personal section, general commentary, subject section and basic skills section.*

The personal section of the profile is filled in weekly by the pupil during one of the daily half hour tutorial sessions. Other parts are filled in by the agreed dates throughout the year with a very full comment being made mid-way through the year. All tutors follow up the mid-year sending home of profiles with home visits and they offer to parents the chance of making appointments to see particular subject staff.

The profile seeks to view assessment as communication and evaluation, with as full a picture as possible being given of the pupil.

Finally, the conference heard from Don Stansbury, Edward VI College, Totnes, Devon, about the Record of Personal Experience which he uses, and from whom further details may be obtained (see **Forum** Vol 21, No 2 *Discussion* Section). Central to the RPE is the development of the individual's personal qualities. The RPE is a personally compiled record of activities, interests and experiences both in and out of school, made during the fourth and fifth years of secondary education. It is not an assessment nor does it produce pupil profile. The pupil has total control over what is written and there is no teacher contribution whatsoever. The Record provides an opportunity for each pupil to record what he thinks is important. It enables them to show initiative and enterprise and provides them with a chance to exercise responsibility. Through recording their experiences they come to know themselves better what their skills are and what they want to do when they leave.

During the afternoon session conference participants joined one of three discussion groups on each of the systems described above and the day ended with an open forum discussion.

It was noted that the take up rate of the Scottish Profile had been disappointing and the RPE was not used to feed information to teachers not involved in its implementation.

The three profile systems were well received, with people generally seeing them as useful adjuncts to the public examinations though not, as yet at any rate, a replacement for them.

DAVID TOMLEY

# A Constructive Response to the APU

---

## Maurice Galton

Maurice Galton is a Senior Lecturer at the University of Leicester School of Education where he is Co-Director of the ORACLE Project which is evaluating different teaching styles in the primary school. He has taught Chemistry in a number of schools and at the University of Leeds. In this article he suggests how private classroom assessment may be made more public.

---

Since the end of the second world war, there has been a tendency for British educators to jump on the latest American band wagon, often at the precise moment in time when the idea or theory was about to be devalued in its country of origin. In the late 50s there was programmed learning, the panacea for more efficient teaching and the answer to the special needs of bright and dull pupils alike. This was followed by a flood of curriculum theories, competency based teacher evaluation and a rapid growth in demands for more complex educational technology at all stages of schooling. The latest of these American exports is, of course, accountability.

There are some striking parallels between the American experience and what is now happening here as the demand for accountability, which in the popular view is seen largely as being concerned with checking up on standards, becomes stronger. One of the American tests, the IOWA tests of basic skills, is now marketed here as a diagnostic instrument under the name of the Richmond Tests. The publishers of these tests are now offering a computer marking facility and it would appear to be a strong favourite among the forty or so local authorities who have now instituted monitoring schemes. The test is advertised as a diagnostic instrument but many local authorities use it as a norm-reference test for making comparisons between pupils and schools.<sup>1</sup>

Although the DES sent representatives from the Assessment of Performance Unit and the National Foundation for Educational Research to study and evaluate the American scene (Burstall and Kay 1978), it would seem that not all the lessons of the American experience have been successfully assimilated. The DES is clearly anxious about the over zealous attitude of some local authorities. Two years ago at the annual conference of the NFER, the then Secretary of State for Education, Mrs Shirley Williams, urged local authority representatives present to take a more cautious approach to monitoring and warned against over-reliance on standardised tests. She advised them to wait the development of new materials currently being produced by the NFER for the APU. Like its American counterpart, NAEP (National Assessment of Educational Progress), the APU will use a technique called 'matrix sampling' where each sample of pupils is tested on a random sample of items. It is claimed that this technique, apart from the advantage of not burdening any one pupil or any one school with excess testing, also diminishes the impact on the curriculum of a large scale assessment programme. The first APU survey in mathematical performance took place during May 1978 with just under a fifth of all primary schools involved. With any one school, only one third of

eleven year old pupils were tested, not all of them doing the same test.

As with the American national monitoring programme, the APU, advised by the NFER, also wishes to move away from norm-reference testing. The procedure they favour is based on the construction of a series of item banks (Choppin 1978). The banks will contain items designed to assess a large number of objectives covering a wide range of content and taking into account different approaches to the subject. Thus in theory a school will retain complete control over what it decides to teach and the methods it uses. Teachers, for example, would still be free to choose either a traditional or a modern approach to mathematics. Items will then be selected from the bank to meet specific curricula requirements but every item will have been calibrated so that they measure a known level of performance in the specific skill listed under the statement of objectives. This kind of development requires a different psychometric approach from the conventional item analysis employed when constructing the norm-referenced tests.

The National Foundation for Education Research are developing these item banks using a technique known as RASCH Analysis (Willmott and Fowles 1974). The theory of this technique and its limitations are discussed and criticised elsewhere in this issue by Harvey Goldstein. There is here, however, a striking parallel with the American situation in that just as NAEP were unable to evaluate adequately their approach, so the first survey of performance for the APU has already taken place before any clear consensus exists about the validity of the RASCH model. It appears that in America the political consequences of admitting that the tests 'do not test what they were supposed to test' were too damaging in view of the expensive nature of the operation. Rather than abandon the exercise the testers have fallen back on a norm-referenced approach. If the same thing were to happen to the results produced for the APU by the NFER, then the best that might be expected would be for those responsible for the survey to argue for caution when interpreting the results, since the future of the Foundation is very much bound up with these new developments in testing. Such reservations are likely to be forgotten in the ensuing debate by politicians and vested interest groups seeking to exploit the results for their own advantage.

The position of the NFER is indeed interesting and also slightly disturbing. Much of the funding currently going to that body is from the DES to support the activities of the examination units. The NFER are responsible both for the mathematical and language development programmes of the APU. Only the Science development has been handed

out to research teams in the universities. Personal and social development, aesthetic and physical development are still in the exploratory stage but the Foundation will no doubt be making its bids in these areas too. For sometime the DES has expressed anxiety about the way that some local authorities have 'jumped the gun' on monitoring. It is hoped that local authorities will be guided by the experience of the Foundation while working on the APU material and, if the item banks are to prove viable, the materials will have to be sold to local education authorities as well as to the APU. In the 1980s therefore the NFER will have a virtual monopoly of the testing and assessment procedures used within the school outside the traditional GCE and CSE examination pattern. With so much at stake, such a concentration of power cannot be a good thing for the schools where in the past the partnership between central and local government, in cooperation with the teachers, has ensured a system of checks and balances which prevented domination by any one vested interest group. The question therefore arises as to what teachers might themselves do to restore this balance of power.

A realistic starting point might be to accept that some form of accountability including monitoring aspects of pupil performance is now with us and that during a period of falling roles and teacher unemployment it would be difficult to mount support for a widespread campaign of opposition to the APU. There is no reason to believe that the situation in Britain would be any different to that in America where views tend to be somewhat polarised and many teachers, in states where accountability laws operate, express a measure of satisfaction at knowing exactly what is demanded of them in terms of their pupils' performance.

Both Bennett (1976) and Barker Lunn (1970) showed that there was a sizable number of primary teachers who still carried out regular testing. It is this worry that some teachers will settle for a minimum and only teach what is available in the item banks that overshadows the APU's emphasis on the positive aspects of monitoring experiment. It is claimed that the information collected will lead to more informed decision making about allocation of resources at both national and local level and that this is as important as comparing standards. But it is the use of these testing procedures by the local authorities that poses the more serious threat to the teachers' autonomy. In Hillingdon, for example, the NFER cooperated in preparing test materials to aid the transition from primary to secondary school (Summer 1977b). The purpose of the exercise was to provide information about pupils and courses to the secondary schools so that they could plan their first year curriculum more efficiently while allowing primary teachers to retain as much freedom as possible over what to teach. In fact there has been no attempt to evaluate the use of these new pupil performance profiles by the secondary schools and at the moment it has proved impossible to do so because the introduction of the new procedures coincided with the transition to comprehensive education, which of itself necessitated considerable upheavals and rethinking of courses. Instead what has happened is that these new assessment procedures have been followed by a lengthy list of objectives designed to mirror the areas which the tests cover. These objectives are called 'guidelines' and look remarkably similar to some CSE examination syllabuses. Thus the continuity between the primary and secondary sectors will be achieved not by any adjustment in the first year at secondary level but a gradual imposition of uniformity at the primary stage.

If *confrontation* is rejected but total acquiescence or *capitulation* deplored, what then remains? Although, in theory, teachers will be asked to write items for these new banks the technical expertise needed for the analysis of such data is always likely to be beyond the grasp of all but a few experts. Thus in any such *cooperative* activity the partnership will be very one sided and control of what goes in or out of the bank is likely to rest with the researchers rather than the teachers. In any case it is clear that many of the objectives which teachers have are concerned not with behavioural outcomes as such but with the processes of learning. This is recognised, to some extent, by the work of the APU's group concerned with Science, where skills of *observation*, *inferring* and *designing experiments* are to be assessed. Progress reports, however, are not encouraging and there are said to be signs of strain within the centres at which this work is being developed.

If there is to be an important element of teacher control over pupils' assessment then the subjective measures, commonly used by teachers as part of normal classroom practice, must be refined and made more explicit, so that they can be seen as clear cut, legitimate, alternative procedures. In one of the early NFER studies teachers were told that their pupils had only progressed during the year by one and a half WITS and were thus one WIT down on the expected norm.<sup>2</sup> Most teachers in that group apparently refused to accept this finding, insisting that *they knew* their pupils had learned more. The problem was that they had no public evidence with which to support this counter-claim. It is precisely this kind of problem that the new 'Teacher-based Assessment' procedures are designed to overcome.

There seems to be a general pattern to the development of this kind of approach illustrated by the work of Wynne Harlen at Reading (Harlen et al 1977), Anne Jasman and the ORACLE project at Leicester (Jasman 1978) and Harry Black in Edinburgh (SCRE 1977). Typically teachers are brought together and asked to state what skills they expect pupils to use during the course of a lesson. They are asked to indicate what kinds of evidence would be acceptable in support of these expectations. Gradually, through discussion, a consensus emerges within the group and different criteria can then be listed and classified according to some hierarchical principle. Checklists are then developed and pupils rated after observation. A pupil's performance profile is gradually built up over time and can be used as a diagnostic instrument either to indicate where additional help is needed or to allow teachers to plan future work activities. Within one group of teachers developing the procedures this seems to work well. After an initial period of difficulty the group members usually become very enthusiastic. It is claimed that it brings a greater awareness of teaching and there is enough evidence to suggest that these assessments are reliable, in respect that different teachers agree very well in their assessments of the same pupils in a team-taught situation. The process by which the teachers acquire such skills tends to be somewhat drawn out. Somewhere between one and two years can be taken up with teachers meeting at regular intervals, taking materials into the schools and trying them out on the pupils.

Clearly teachers who are prepared to involve themselves in this way are likely to be both concerned and enthusiastic. With so much involved, such techniques are not likely to appeal to those who are prepared to settle for the lowest common denominator and to accept the local authorities

'guidelines' as a performance specification. In all three studies mentioned above the attempts to persuade a wider audience to take the materials developed by the teachers groups has not so far met with much success.<sup>3</sup> Since learning how to use the checklists takes time many in the profession feel that this can be better spent doing other things with their pupils. If testing must be carried out, they argue better to use a conventional assessment programme which takes up one morning from the school term rather than requiring frequent monitoring of individual pupils.

At the heart of the problem lies the fact that these new assessment procedures demand a different style of teaching from the one conventionally used. A number of observation studies have shown that the didactic expository style still finds favour with a majority of teachers both at junior and secondary level (Eggleston et al 1976; Galton et al 1980). Regular monitoring of a pupil's performance in terms of the processes of learning rather than carrying out intermittent assessments of the final product makes totally different demands on the teacher. She needs time to engage in a dialogue with her pupils, and must be able to observe them while they are working cooperatively together.

Yet the ORACLE studies show that the amount of teacher and pupil contact at this intimate individual level is very slight, that collaborative group activity involving pupils sharing ideas almost non-existent and that many teachers are forced to spend large amounts of their time marking silently. In just over half the classes studied forty per cent of the pupils who were observed spent up to the equivalent of one day a week either waiting for the teacher or involved in a wide range of distracting activities. Yet the current suggestion of the inspectorate is to increase the class sizes so as to free some teachers who could deal with pupils having special difficulties. It is argued that if you have a class of 35, four more will not make much difference. It remains to be seen whether the teacher unions can persuade the politicians that falling roles should lead to decreasing class sizes rather than school closures.

Despite these difficulties there seems no real alternative but to pursue a vigorous policy of in-service training in these new assessment and evaluation techniques at as many levels as possible. While the traditional 'one day' course can probably do little to encourage the use of monitoring inside the classroom the continuing increase in part-time higher degree students provides a chance for teachers to develop an understanding of the skills involved. This training could provide a nucleus of personnel to run school-based in-service courses, the ideal environment for training in teacher-based assessment techniques. Courses run by teachers for their colleagues start with a natural advantage, given the genuine fears which may exist over local authority attitudes to accountability. Perhaps the role of the head in stimulating school based activities needs also to be looked at.

In the past researchers developing these new assessment techniques have tended to be over ambitious. Both the Scottish 'Pupils in Profile' and Wynne Harlen's 'Progress in Learning Science' checklists are far too comprehensive and elaborate. Both look impracticable and time consuming so that natural resistance to using the checklist is built in from the start. The importance of the method is not what is being assessed but the technique by which these assessments are carried out and teachers need first to be persuaded to move to a less directed style of teaching if reliable assessments are to be made. It is therefore just as valuable to train someone to assess one small element in

the curriculum rather than trying to sell a 'complete system'. At the outset it will be necessary to 'cut corners' until more people have gained confidence in using these techniques. In the 'ORACLE' study, for example, structured exercises have been provided which mirror the skills being assessed by teachers through observation. These structured exercises, not only provide clues as to what the teacher should look for, but also provide a means by which the teacher can compare her judgement against other kinds of evidence. In principle, a similar use could be made for the item banks currently being developed. In many cases a teacher's judgment and the more formal assessment will agree. The teacher is then free to concentrate on observing, in greater details, those cases where a mismatch occurred. This might go some way towards meeting the problems of using these techniques in large size classes.

Given support from those responsible for reflecting officially teacher opinion and a commitment from those responsible for in-service training, teacher-based assessment could provide a satisfactory alternative in meeting the demands for more 'objective' information about pupil attainment. The great value of promoting this exercise would be that, at last, it would begin to erode from the public mind the assumption that the old fashioned 'traditional' forms of teaching go hand-in-hand with strict accounting. If this were to happen then we might consider re-exporting accountability theory to America in a healthier and more valuable form.

#### Notes

- 1 A norm-referenced test compares a pupil's performance in relation to the other pupils taking the test. It is designed to spread pupils out as much as possible in terms of their test scores giving the customary normal shaped distribution. A criterion-referenced test is merely concerned about whether a pupil has achieved the required standard of performance. Such distributions are bimodal consisting of a group who have and another who haven't mastered the set tasks. For a programme of national monitoring, criterion-referenced tests are obviously more appropriate. However, there are difficulties in constructing and analysing such tests (see Sumner and Robertson 1977a). In America, researchers have got as far as specifying the criteria but in the absence of an agreed procedure for item analysis they have been forced to use techniques which are more suitable for a norm-referenced approach. Tests of this kind have come to be known as 'objective referenced' and represent a rather unsatisfactory compromise between criterion and norm-referenced testing.
- 2 Under the RASCH model, the chances of a pupil making a correct response to an item depend only on the pupil's ability and the item difficulty, both measured in the same units. The unit used by the NFER is called the WIT. It is so chosen that when a person's ability exceeds the difficulty of the item by five WITS then there is seventy-five per cent chance that the item will be answered correctly. As one might imagine, there are also half-wits, and according to the NFER the average pupil is likely to make two and a half wits progress per year of schooling.
- 3 In a recent example, as part of the ORACLE study, teachers were asked about their assessment of project work. The majority registered extreme concern about the judgments they made in this area and when shown a checklist, developed by a group of local teachers, were very supportive and claimed that it was 'just what was required'. Yet after a term not one teacher in the sample had made use of it to monitor their pupils' performance. When teachers on a related in-service course volunteered to persuade their colleagues to try out the checklist on one or two pupils they met with no greater degree of success.

*See page 18 for a list of references*

# The Case for School Self-Evaluation

John Elliott

John Elliott is Tutor in Curriculum Studies at the Cambridge Institute of Education where he is also directing an SSRC sponsored project on Accountability in Secondary Schools. Formerly a secondary school teacher, he worked on a Schools Council Humanities Curriculum Project and was Director of the Ford T Project and Lecturer in Applied Research in Education at the University of East Anglia.

Like most innovatory ideas the concept of 'school self-evaluation' is vague and ambiguous. It can be understood simply as an answer to the question 'Who should accept responsibility for evaluation' regardless of what is to be evaluated. In which case it will pick out the school as an evaluation agency in contrast to other possible agencies like the local or national inspectorate. In this sense one could have self-evaluations of University INSET courses, Teachers' Centres, LEAs, Schools Council Projects, etc. I suspect that like me a number of people understand the idea in a stronger sense than this. I would define it as *the school accepting responsibility for evaluating itself*. Here 'the school' is both the agent and the object of the evaluation. In this article I shall be concerned with 'school self-evaluation' in its stronger rather than weaker sense.

I am going to set out three kinds of grounds on which a case for a school self-evaluation could be argued. The form an evaluation activity takes in practice largely depends on one's reasons for engaging in it. By setting out some alternative reasons for a school engaging in a self-evaluation I hope to indicate the different forms such an evaluation might take in practice.

The case for a school self-evaluation can be argued on at least three kinds of grounds which I will call those of *social*, *economic*, and *professional accountability*.

Schools might be described as *socially accountable* to the extent our society delegates to them the task of preparing children for socially valued roles and tasks. This act of trust places schools under an obligation to provide society, particularly at the level of the local community, with an honest account of its policies in the light of relevant social criteria. For example, a school might present to its governors an honest account of the extent to which its option policy provides girls with equal opportunities to boys for choosing science as a career, or it might present to its parents an account of its curriculum policy for mathematics, in the light of the question 'Is the maths taught in the school the kind of mathematics children will need as members of society?' Within a context of social accountability 'school self-evaluation' largely consists of reviewing curriculum policies in terms of the extent to which they make adequate provision for children's future social needs.

Schools might be described as *economically accountable* if they are under an obligation to demonstrate to those responsible for allocating resources at central and local government levels that they are giving *value for money*. In order to establish a system which makes schools economically accountable to the providers of resources the latter have to get them to see themselves as factories manufacturing clearly defined and measurable outputs in the form

of 'learning outcomes'. Once these out-puts have been specified schools can be required to justify the resources allocated to them by demonstrating that they are achieving an acceptable level of productivity ie in relation to the resources available. Such a demonstration will take the form of an evaluation of the extent to which pupil performance matches the appropriate level of out-put.

At this point it may be useful to spell out the essential differences between 'school self-evaluations' based on economic accountability and those based on social accountability. First, there is a difference in focus. An evaluation in a social accountability context will focus on the curriculum policies of a school; the courses of action it undertakes with respect to the selection of content and its organisation for purposes of learning. An evaluation in an economic accountability context will focus on the performance of pupils. Secondly, there is a difference in criteria for appraisal. Social accountability implies a form of evaluation which is concerned with the social acceptability of policies. Economic accountability implies an evaluation concerned with the economic acceptability of pupil performance levels (See MacDonald 1978). Thirdly, this difference in evaluation criteria has methodological implications. Social evaluation essentially involves the use of qualitative methods while economic evaluation involves methods of quantification. The former is more concerned with the quality of what Parlett and Hamilton (1977) call 'the learning milieu' than with the amount of learning achieved.

I have just started interviewing parents with a view to discovering the ways they assess the institutional merit of schools. So far as I have talked to a very small number so must be wary of stating more than a tentative hypothesis. But I am rather surprised to find, even amongst the few parents interviewed, that they are more concerned with the social relevance of their school's curriculum policies than with its productivity in the form of exam results. In evaluating schools they are concerned with the extent to which the school's policies make adequate provision for their children to learn what is socially valuable. This is not to say that they regard quantifiable information about pupil performance as unimportant. But they appear not to see such information as central in evaluating the work of a school. This explained for me another surprise I had when interviewing: I have found few parents wanting to see schools publishing their exam results to the world at large. They don't want to see 'league tables' of results published, feeling that such out-put information gave a distorted view of the relative merits of schools. The parents I am talking to are sufficiently realistic to appreciate the fact that their children, rather than their school alone, must bear some

responsibility for the amount they learn. They are prepared to hold the school responsible in a negative sense, when its policies and practices appear to prevent children from realising their full potential. But in the absence of this negative evidence they are sensible enough not to blame the school for learning failures.

Fourthly, economic differs from social evaluation with respect to its audience. A school that sees itself as economically accountable will tend to evaluate itself for its resource providers, namely, its LEA. The direct line of accountability is between the school and the LEA bureaucracy. On the other hand a school that sees itself as socially accountable will tend to evaluate itself for the local community, as represented perhaps by its governors, and its parents.

At the present time some LEAs are attempting to get schools to see themselves as economically accountable units. The indicators are all there; advisers encouraging schools to specify their aims in terms of pupil achievement criteria; in-service courses on teacher-based assessment of pupil performance; the bureaucratic processing of examination and test results submitted by schools. Even at national government level some HMIs have been encouraging schools to adopt the achievement criteria developed by the APU as a basis for school self-evaluation (see Kay 1977). Meanwhile, at the level of the local community, some schools are breaking new ground in presenting school governors and parents with detailed information about curriculum policies as a basis for discussion.

Are these two emergent kinds of 'school self evaluation' compatible with each other? Personally I think not. A school which comes to see itself as accountable for pre-specified out-puts will ultimately allow economic ends — the cutting of costs — to distort the social value of its curriculum policies. They will become less responsive to a variety of social needs.

Determining the social quality of curriculum policies is a difficult and complex affair. Any particular policy will carry a variety of social advantages and disadvantages which are not always easy to detect and weigh up. Moreover, the balance of advantage to disadvantage will vary over time. What looks like on balance an undesirable policy in the short term can turn into a desirable one in the long term, and vice versa. It is not only important to scrutinise the intrinsic qualities of a policy but also its short and long term consequences, including unintended side-effects. Over time its qualities shift and change. The social evaluation of curriculum policies is therefore a matter of piece-meal ongoing social criticism. There can be no absolutely conclusive summative evaluation. There will always be times when the school self-evaluation suggests that on balance *existing* policies need to be changed and a decision ought to be made. But even at the point of responsible decision one can never be completely sure.

Economically based school self-evaluation, emphasising as it does, quantity rather than quality, will tend to neglect and ignore the social quality of learning. In taking pre-specified learning outcomes as the criteria for an evaluation of itself the school comes to view the value of its policies in strictly instrumental terms. They are desirable if they maximise intended learning outcomes, undesirable if they do not. This kind of goal-orientated evaluation blinkers schools to both the intrinsic social value of policies and to their manifold social side-effects. As Scriven has pointed out a policy may be a good one even when technically it is a failure, simply because its social side-effects bring benefits

which far outweigh its failure at goal-achievement (1977). This is no argument against goal-setting but a social accountability argument against evaluating exclusively on terms of them. Social accountability requires a willingness in the part of schools to change and modify goals in the light of changing conceptions of the social qualities of their policies. Economic accountability fixes and masks these qualities and makes the school socially unresponsive.

The emphasis on quantity rather than quality also carries another distorting effect on the social value of policy. Certain learnings are easier to measure than others. Those that are not will tend to get neglected irrespective of their social value. A classic example is provided by the APU's pamphlet entitled **Language Performance** (1977). It states 'There are four modes of language — listening, speaking, reading, writing.' It then goes on 'Following the advice of the Bullock Committee, the language group has started with writing and reading. The other two modes are important and all four are interconnected, but listening and speaking are even more difficult to assess adequately.' If schools then come to accept the kind of performance criteria defined by the APU as a basis for self-evaluation they may find themselves focusing on the more easily quantifiable aspects of language with a consequent distortion in the overall quality of their language policies. It is certainly arguable that the social importance of speaking and listening is greater than that of reading and writing.

## Professional Accountability

School can be described as *professionally accountable* if they are under an obligation to demonstrate to the teaching profession that their policies protect and foster educational values. In addition to interviewing parents about school accountability I am also interviewing teachers. The phrase 'a professional attitude' occurs again and again. When asked to explain what they mean teachers tend to talk about 'a concern for the needs of children'. These needs refer not so much to social as to personal needs. The educative responsibilities of the teacher as opposed to his social responsibilities are towards children as potential persons rather than as potential occupants of social roles. Of course, a child cannot develop as a person without developing in a certain role. But he can certainly occupy a role, even competently, and remain as a person alienated and undeveloped. Children develop as persons when they learn to autonomously choose their roles for the intrinsic value and interest they discover in their performance rather than for extrinsic reasons. If the school curriculum is the medium through which social roles are made available to the child its teaching methods are the medium through which he is related to them. The teaching policies of a school can enable or constrain the development of personal autonomy. They can educative or miseducative regardless of the social value of the curriculum they transmit.

A school self-evaluation in a context of professional accountability will focus on either common teaching policies or those for which the school delegates responsibility to individuals or groups of teachers. The criteria employed will be related to the extent to which classroom teaching enables and protects the development of personal autonomy in the learning situation. The audience will primarily be the teaching profession and in particular the staff of the school concerned.

The different forms of school self-evaluation entailed by social, economic, and professional accountability can be summarised as follows:

Self-Evaluation	Social Accounting	Economic Accounting	Professional Accounting
Focus	Curriculum Policies	Pupil Performance	Teaching Methods
Criteria	Social Acceptability	Economic Acceptability /value for money	Educational Acceptability /enabling personal autonomy
Methodology	Qualitative	Quantitative	Qualitative
Audience	Local Community/ Governors	Local Government /its officers	Teachers /The School's staff

In my view socially and professionally based school self-evaluations are not necessarily incompatible and must eventually 'marry' if schools are to avoid neglecting the educational needs of children in favour of their social needs, and vice versa. School self-evaluation in a context of economic accountability necessarily has a distorting effect. The educative value of a school's teaching methods lies not in the amount of content they get children to learn but in the way they enable children to learn that content. In a context of economic accountability the educative quality of teaching becomes an irrelevance.

Some would argue that schools cannot be trusted to evaluate (account for) their own policies objectively. Such evaluations will tend to be biased in the direction of presenting the school in a favourable light. Only an external evaluation can satisfy the requirements of objectivity. My reply to such arguments would be that if schools cannot be trusted to evaluate themselves honestly then by implication they cannot be trusted to make responsible decisions.

A school that is reluctant to engage in an honest self-appraisal reveals a lack of genuine concern about the extent to which its policies are socially and educationally worthwhile. Policy decisions in schools can only be made responsibly when they are guided by these concerns.

*Continued from page 27*

For the school, as with its individual children, assessment of all aspects of 'physical, intellectual, social, emotional and aesthetic areas of development' is valid and necessary. A full, regular review is essential in order to identify strengths, weaknesses, particular needs and to explain any fall in levels

*Continued from page 29*

the relationship between the curriculum practices of the school, the principles on which the school operates and the general assumptions made by educationists about the nature of the curriculum process. It includes three or four advisors from outside the school, including one member of the PARB.

What did become obvious at Lorne was the strength of the consensus about the positive aspects of the school. There was an overwhelming re-affirmation of the principles on which the school operates, and of the need for Huntingdale to be both a learning *and* a caring community.

Without the Lorne Seminar, the Planning and Review Board would have been an end point rather than simply one (important) point in a continuing process. One suspects, indeed, that much of the value of the PARB Report would simply have been dissipated in an aura of self-congratulation without the hard-headed follow-up which placed the

We are currently witnessing attempts in some local authorities to externally evaluate schools more systematically. In my view such developments, whether they take the form of externally developed and imposed tests or local inspections, are indicative of a tendency to transfer power over decision-making from the schools. Externally imposed evaluation systems are essentially the means of getting schools to comply with policies decided elsewhere. They are grounded in the knowledge that power can only be exercised over others by constantly monitoring their activities. The weaker the monitoring the less people will comply. The development of external evaluation systems is indicative of the current crisis of confidence in our schools. Although the 'cry of accountability' is used to legitimate it, the truth of the matter is that such a development makes it impossible for schools to exercise any genuine accountability to society or themselves.

Perhaps the advocates of external evaluation are correct in assuming that schools cannot be trusted to formulate responsible policies. If a willingness to provide society or even themselves with honest self-appraisals is a sign of responsible policy-making, then perhaps they are right. The onus now I think is on schools to demonstrate that they can be entrusted with policy-decisions by accepting more responsibility for evaluating themselves.

#### References

- |   |  |
|---|--|
| MacDonald, B (1978)                       | 'Accountability, Standards, and Process of Schooling' in <b>Accountability in Education</b> ed Becher, A and Maclure, S (NFER) |
| Parlett, M and Hamilton, D (1977)         | 'Evaluation as Illumination' in <b>Beyond the Numbers Game</b> ed Hamilton D et al (Macmillan)                                 |
| Kay, B (1977)                             | 'The Assessment of Performance Unit; its task and rationale' in <b>Education 3-13</b> , 4:2, 1976                              |
| Scriven, M (1977)                         | 'Goal Free Evaluation' in <b>Beyond the Numbers Game</b> ed Hamilton D et al. (Macmillan)                                      |
| DES Assessment of Performance Unit (1977) | <b>Language Performance</b> - Information Leaflet.   |

of achievement; to stimulate dialogue between teachers, pupils and parents, and, above all, to ensure that the individual and the school know just where they stand. Since the school must be accountable for standards to its public and its clientele it must be ready to initiate change for the sake of progress.

whole exercise much more firmly into context.

In any short list of implications for the UK must stand consideration of alternatives to the Governing Body and the role of the Inspectorate. Huntingdale's extensive re-appraisal was effective as a spur to development because it was a corporate venture. Changes will be undertaken because the conclusions reached at Lorne were those of the staff, the students, their parents and the school's own Council. Nobody was acting under duress or made to feel they were responding to correction. And in the end, only those working in continuous contact with a school can decide what will and what will not work for it. A review, call it what we will, has to find a way of being both critical and creative: it has both to reveal possibilities hitherto unrealised and to inspire those on the job to devise the strategies appropriate to realising them. We would claim that in these respects, Huntingdale's review system was highly successful.

# Self-assessment at a London Comprehensive

---

## George Varnava

George Varnava is First Deputy Head at Pimlico School. Formerly Head of Modern Languages at Holland Park School, he has been fully involved in mixed ability teaching and comprehensive organization. Here he comments on how one school has responded to the challenge of self-assessment.

---

Two years ago, ILEA's *Keeping The School Under Review* invited schools to undertake a systematic exercise in self-assessment; in the Secondary section, to clarify what it is the school is attempting to do and whether it is satisfied that what it is attempting is right. Those courageous enough to answer its direct, practical questions with complete honesty found, no doubt, clear directions to areas of necessary or desirable improvement.

The Secondary section of the document directs its questions to facts, people, policies and procedure, concluding with an 'acid test': 'Would I recommend a colleague to apply for a post at the school?' and 'Would I recommend the school to friends for their children?' It is tempting to apply the ultimate test: 'Would I send my own children to the school?' A final note on 'Action' is designed to ensure that the exercise is constructive.

This proposed method aims to stimulate the kind of enquiry and discussion necessary to any re-definition of the school's objectives and the general educational principles by which it functions. As the introduction points out, each question is likely to suggest others. Under Resources, for example, a separate count of male and female staff, the number of teachers who have children themselves, length of service, etc., may go some way to explaining staff attitudes to their work and their professional development. Similarly, a count of 'problem' children or one-parent families will show whether or not a shift of resources is necessary. It is worthwhile, too, to take into account any additional use of the school premises and identify any particular architectural features that affect organisation and working conditions: a main hall too small for full assemblies; specialist rooms that cannot be used for general teaching; dark, secret corners that invite delinquency, and as many more examples, surely, as there are schools. Even the school roll deserves careful analysis: at Pimlico, a roll of 1,591 constitutes only 1,156 families.

The entire process of self-assessment, nevertheless, can easily remain a bureaucratic exercise if not conducted with the determined intent to make changes where changes are necessary, in both particular aspects of school life and in standards of efficiency. At Pimlico, the exercise was conducted partly through a full staff conference, neatly entitled the 'The Common Task', aimed at identifying areas of concern and involving staff in solving problems and, subsequently, by a complete revision of the Staff Handbook, a document of over 60 pages covering policy, organisation, resources, curriculum and daily routines — a combination of reference book, standing orders and a statement of intent.

Apart from bringing factual detail up to date, major changes were necessary in the areas of staff responsibilities — following the evolution of the role of the Form Tutor and a general shift of responsibility for classroom discipline from pastoral to academic staff; in regulations regarding security in the light of the Safety at Work Act; in school uniform — resulting from prolonged discussion involving staff, School Council and Parents' Association, and finally, in the areas of the assessment of pupils where both mixed ability grouping and setting by ability occur. This last change, subject of a working-party's discussions lasting almost two years, is perhaps the focal point in any exercise of self-appraisal, for it is ultimately the criteria by which children are judged that determine the school's philosophy and objectives, its policies and organisation, and its performance.

Here is Pimlico's instruction to staff:

### Assessment (Grades and reports)

#### *General*

Continuous assessment of the individual pupil is made by awarding two grades for each subject. These are recorded in the Form Grades Book prepared by the Form Tutor, and on twice-yearly reports. Each child is awarded a double grade in each subject on a 5 point scale, the first digit being for the quality of the work (attainment) the second for effort.

1 — for excellent; 2 — good; 3 — average; 4 — below average; 5 — poor. For example, a child who is given in English an assessment of 2.1 is judged to have produced work of good quality and have worked very hard to achieve it. The continuous assessment of attainment and effort provides the school record of the individual pupil's academic progress throughout his school career.

Progress may be seen as competition against oneself and this simple, regular form of assessment aims to give each pupil the incentive to compete against his last performance. A combination of both mixed-ability grouping and setting occurs within certain departments and within years. To ensure, therefore, a correct scheme of assessment recording, involving the use of a letter code which precedes the grades, is used throughout the school from years 1 to 5.

#### *Procedure for assessment*

GRADES — attainment )  
                                  — effort           ) both on 5 point scale.

awarded for all pupils and recorded in grade books.

Grades preceded by letter A, B, C or M

A, B or C denoting level of group as fixed department;  
M denoting mixed-ability.

(N.B. A, B, C M *not* used for 6th)

REPORTS

1st/2nd Years:

- a Effort grade only, preceded by M (N.B. A, B or C for Latin or sets in 2nd Year);
- b Exam mark where appropriate
- c Comment (explanatory and advisory)

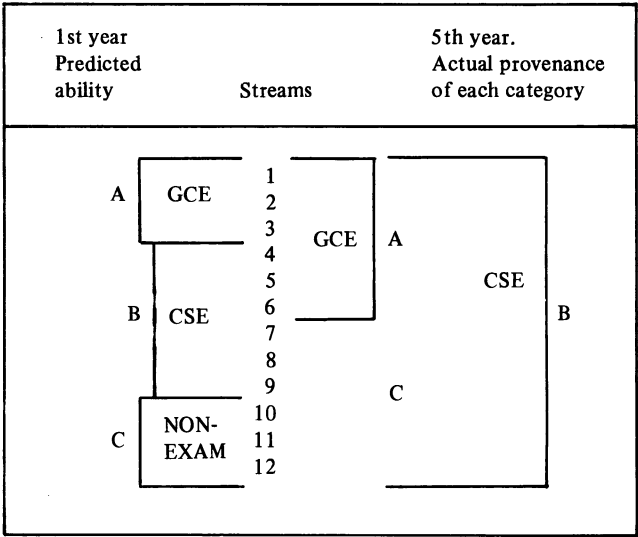
3rd/4th/5th Years:

- a Both grades, preceded by A, B, C or M;
- b Exam mark where appropriate;
- c Comment (explanatory and advisory)

6th Year:

- a Both grades (no A, B, C or M);
- b Exam mark where appropriate;
- c Comment (explanatory and advisory)

Here, an attempt has been made to reconcile the fundamentally different approaches in assessing children in mixed-ability groups and those set by ability. In mixed-ability groups, co-operation between pupils is more appropriate to the learning situation than the encouragement of keen competition. Any objective assessment of ability, therefore, should not hinder the motivation of the individual pupil. In sets, assessment of attainment is progressively related to potential in public examinations. A conventional statistical approach to examination results normally gives no more than a crude picture of a school's performance. Marten Shipman's very useful 'The Presentation of Examination Results' in *Contact* (24 Nov 1978) indicates how important it is to assess performance in terms of the school's intake characteristics and its examination policy. The following figure shows how a school's achievement might be interpreted if the question asked is not 'how many passed?' but 'who passed?' — a question, surely, that is of much greater interest to those directly concerned.



Parallels between the assessment of individual children and of the school as a whole are obvious, whether the assessment is made by comparison with others or with previous performance. In the Secondary section of *Keeping the School Under Review* two brief questions only refer to the assessment of children: H5 (Arrangements for learning) 'How do we assess the quality and quantity of work produced?' and I (Departmental Self-assessment) 'What steps do we take to assess and record progress?' The Primary section gives greater importance to 'Attainment' with fuller, more searching questions.

*Continued on page 25*

# A Year of Evaluation

A R Delves and John Watts

The Principal of Huntingdale Technical School, Melbourne, Australia and John Watts, Principal of Countesthorpe College in Leicestershire have collaborated in presenting this report of the evaluation process in which both were involved at Huntingdale.

The conventional British approach to school assessment is patrician. The Full Inspection requires a team of Her Majesty's Inspectors to investigate all aspects of the school, to go away and after due deliberation to present a confidential report to the Governors of the School and the Secretary of State. In practice there is likely to be a lot of valuable discussion between individual Inspectors and classroom teachers, but in principle this is not part of the process. As the reporting Inspector told the staff at Countesthorpe in advance of their Full Inspection of 1973, the Inspectorate cannot at one and the same time report and enter into discussion.

It might be thought that such procedures ensure objectivity, free from favouritism. Unfortunately, that view turns on the assumption that in every aspect of schooling, from organisation to pedagogy, from examining to counselling, the Inspectorate knows best and the teachers will be judged on the extent to which they have 'got it right'.

Indeed the desire for accountability has often led politicians to call for a local Inspectorate that would check whether a school was doing things 'the right way', on the analogy of industrial safety inspectors who can check conformity to regulations. It is of course a vain hope.

There are no regulations that add up to a prescription of how a school should be run, and in our changing times, there is no certainty that a team of Her Majesty's Inspectors have got all the answers for the teachers. Sometimes things will have moved on to a point where the HMIs have not even got the questions.

Many HMIs have acknowledged a change in role from that of judge-and-jury to one of counsellor, guide and friend. Many must prefer to sit down with the teachers and jointly tease out a problem together. Unfortunately, their own machinery for inspection and report has not undergone comparable re-designing. What schools desperately need is help in the techniques of self-evaluation and consultation, a totally different mode of giving account of themselves.

Teachers are amateurs in self-evaluation where they should be professionals, able to carry out mutual assessments and able to teach the techniques to their students. And to avoid complacency and insularity, they need to learn how to seek and take external counsel. To accomplish this may entail serious revision of our present procedures and of the function of the Governing Body. That it would be worthwhile, may be seen from the experience of Huntingdale Technical School in its evaluation of 1978.

In Australia, government primary and secondary education is controlled by the six states, with the Commonwealth government assuming direct control only in the various

Territories. In the State of Victoria, school assessment was, as in Britain, the province of Boards of Her Majesty's Inspectors who reported their findings to the State Education Department and to the school, although without the element of confidentiality which surrounds British practice. Some eight or ten years ago this situation began to alter dramatically as a result of a series of fundamental changes. At the secondary level there was a significant move towards school-based decision making in a number of areas and this inevitably led to a great diversity in curriculum across that state. At the same time, the teacher unions adopted strong policies of opposition to the system of inspection, chiefly because the promotion of teachers depended almost entirely on inspectorial reports. In addition, the Inspectors, finding themselves often unwelcome in — and even locked out of — schools, sought for themselves the role of 'counsellor, guide and friend'.

These changes greatly altered the quest for accountability of schools. In the secondary high schools, the Education Department instituted a process of School Review Boards. Schools seeking assessment of their work ask for such a Board, which is then set up as a joint venture between the school and the Education Department. Control of this process, however, remains firmly in the hands of the latter. These Boards report to the school and the Department, as did the Inspectorial Boards of old, and there has been, to put it mildly, a mixed reaction to their work.

A further important recent change by the Victorian Government has been in the field of school government. In 1975, the Education (School Councils) Act significantly altered the role of the advisory Councils which had operated in all schools for some years. Firstly, each school community — parents, staff, students and interested people from the neighbourhood — determined the composition of its own Council. (In Victoria, School Councils have never been *political* bodies in the way School Boards are in the United States, or Boards of Governors in Britain).

The powers of the Council are wide, and include responsibility for buildings and grounds, for cleaning and maintenance, for all finance coming in to the school, for the employment (and payment) of non-teaching staff and for ensuring maximum community use of the school and its facilities. Councils may enter into contracts (for example, for building extension) on behalf of the school. Each Council is asked to *advise* (but may not *direct*) the Principal and staff on curriculum, and it has little or no role to play in the appointment of teaching staff who are placed by the Education Department.

In the case of the secondary technical schools, however, the Councils have by law a strong advisory role to play in the appointment of senior staff. In addition, the technical schools have had, by long tradition, control over all their operating finances except for teacher salaries. The importance of this power may be gauged by the fact that Huntingdale Technical School Council controls funds in excess of 150,000 pounds per annum.

The composition of The Huntingdale Council also gives a clear indication of the way in which the major groups in the school community is reflected in the overall government of the school. There are four elected parents, four elected staff, four elected students, four members nominated by a voluntary community group — the school is used extensively by the community at large — one representative of the local government authority and one of the Education Department, the principal, and up to four others co-opted by the elected and nominated members. It should also be

noted that the term 'staff' is used in the school to include all employees at the school, both teaching and non-teaching.

When Huntingdale began in February 1972, the Education Department asked the school to examine alternatives for curriculum facilities and the development of community relationships. In return for this chance to innovate across a broad spectrum, the school was also asked to conduct an assessment of its work after the first five or six years of operation.

## Self evaluation

The process of evaluation began in September 1977, when individuals and groups within the school began to look at the positive and negative aspects of their work thus far. The School Council, at the same time, set up an *ad hoc* Committee to oversee the whole process, taking the view that it was the Council's special responsibility, acting on behalf of the whole school community and of the Education Department.

By early 1978, the criteria for the evaluation had been determined and agreed upon, and it was broadly decided that there was a need not merely to review what the school had already done but also to plan for the next five to ten years.

The evaluation included a range of activities, beginning with the internal evaluation mentioned above. The school also commissioned a 'Community Needs and Response Survey' which was conducted by staff and students of a College of Advanced Education, assisted by Huntingdale staff and students. Students who left the school in 1976 and 1977 were followed up to look at their job or tertiary education situations. A survey of the parents of Year 6 students in feeder primary schools was taken to establish why they would choose (or reject) Huntingdale for their children in 1979.

The School Council conducted an evaluation of its own work late in 1977, as did the Community Involvement Group which had an important role both as a committee of Council and as the management group for the community use of the school's facilities.

The major documents which had been written in and about the school since its beginning were also collected and published in a large volume entitled 'The Evaluation Book'. The Council also commissioned a Planning and Review Board of seven well-known educationists, including two from the Department of Education, to come and assess the school for a week at the beginning of August 1978. There was again emphasis on forward planning as well as on review. The Board was financed from school funds, supported by special Education Department grants, and its task was to report to the school community and the Council, and thence to the Education Department and the Minister of Education. In a very real sense, this procedure turned completely upside down the traditional school accountability process.

Council determined that the composition of the Board had to be such that the members were widely regarded in Australian education with a clear reputation for independence of spirit. At least one member of the Board was, at the very least, uneasy about the school and its work when the Review began.

This was not, however, to be the end of the evaluation, even if it was in a way the high point. The climax was a three day live-in conference for members of the school

community at a government-owned residence some 150 km from Melbourne. At this conference (The 'Lorne Seminar') *all* of the evidence of the various components of the evaluation (including the Review) would be considered and planning for the next five to ten years would take place.

In July 1978, the seven members of the Planning and Review Board (PARB) came together under the chairmanship of Ray McCulloch, Associate Professor of Education, Monash University, and chairman of the School Council of Huntingdale Technical School. The six other members were Jean Blackburn (Australian Schools Commission), John Mayfield (Director of Educational Facilities, South Australia), Barry Fitzgerald (Head of the School of Education, Ballarat College), Bill Johnson and Noel Watkins (both Assistant Directors of Technical Education, Victoria) and John Watts (Principal, Countesthorpe College). As their backgrounds show, they represent a broad coverage of the educational fields in which Huntingdale operates.

The terms given to the PARB by its chairman made clear that it was to be a part of the evaluation process and an aid to forward planning over the following years. The PARB was to spend a week in the school and to present a critical appraisal of the work of the school. All members of the Board were familiar with the school from earlier visits (even John Watts had spent the previous week in the school), and each of them had studied the recently compiled Evaluation Book, the extensive documentation of the school's six years of existence. Some members of the Board had had close contact with the school ever since its inception, and one member, Noel Watkins, had already seen his own daughter pass through the school.

It was made clear to the Board by its chairman that we would be reporting in the first instance to 'the school community'. The report would be public and it would be immediate, delivered on the last day of our visitation. It was, however, also clear that there was a wider interest and that our report would be presented to the Minister of Education for Victoria State.

In spite of initial announcements made to the school there was a tangible suspicion on the part of some staff, that the PARB had been convened to assess individual teacher performance. To establish good faith therefore, the Board announced on arrival that all its deliberations would be open, and that we would give an interim report at the end of each day. The PARB members were introduced to staff on the Sunday evening, and to the whole school on the Monday morning.

## Open dialogue

The daily report-back-sessions, starting on the Monday after school, took an increasing significance. All through the day members of the Board, singly or in pairs, would accompany staff and students, listening, questioning, and then at 3.30, would each comment on their experience, posing new questions, inviting comment, announcing their intentions for the following day. Once it was seen on Monday that the Board was being as good as its word, that an open dialogue was being created, the attendance grew from some forty or so to the two or three hundred who packed the long session of the final Friday afternoon.

The only session that became limited to members of the Board was that of the Thursday night, during which the final reports were hammered out, agreed and drafted. Not that the meeting was exclusive: very pertinent contribu-

tions were made by the Principal, who sat it out, and the Caretaker, whose perceptive and optimistic summary heartened us all round about midnight.

The PARB was given access to all records and personnel, but this access was reciprocal (in contrast to a British Full Inspection which is not open to reports or comments from parents or governors). Between them, members of the PARB saw teaching in progress, students in unsupervised study, pastoral meetings, staff meetings, curriculum discussions, a meeting of the School Council, and adult students; they examined equipment and resources, explored the site and the new buildings under construction.

The PARB had agreed initially that it would try to consider the ways in which the school saw itself serving its community, and the feasibility of the proposed use of its resources over the next five to ten years. The Board also felt it should comment on the ways in which Huntingdale's unique features might or might not be contributing any useful model for wider application. The task was therefore no mean one, and well comparable to a Full Inspection by HMI.

The climax of the week was Friday afternoon's presentation. Each member of the Board read a report, agreed in draft the night before, each covering one particular aspect so as to make a coherent whole. The presentation was public, and subsequently edited into a final document of some forty pages. Probably what is more significant procedurally is that what the Board had to say collectively was delivered before departure. There was no prolonged deliberation behind closed doors and no final anonymous and confidential document delivered to the Principal and Governors. It is also significant that the conclusion of this intensely busy and stressful week came *after* the PARB's delivery and all the votes of thanks, with a satirical masque performed by students and a celebratory buffet supper for everybody. They may well have been celebrating the PARB's summary statement which concluded by saying that their school 'already ranks among the most prominent and promising break-throughs in the present world-wide quest for new and more appropriate forms of popular education. . .'

111 members of the school community attended the Lorne Seminar for three days in November. Seventy-one of those were staff, twenty-two students and eighteen parents and other interested people. Sixteen of the total were also members of the School Council.

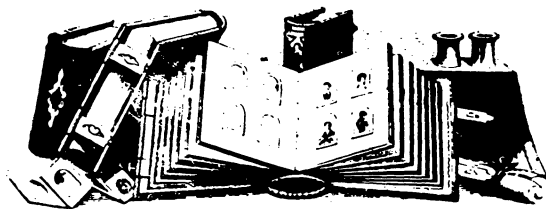
Prior to the Seminar, agenda items were proposed by anyone who wanted to put something forward, and these were distilled by the Agenda Committee into five main Topics: (a) A Caring Community; (b) Curriculum; (c) Beyond a Secondary School; (d) School and Community; and (e) HTS 1984.

Each Topic was introduced by a 'lead' speaker who helped to set the parameters for the subsequent discussion which took place in groups of about eight to ten people. At general sessions motions previously committed and new motions proposed by groups were debated and adopted only if a broad consensus was reached. Over forty resolutions were approved by this process.

One important outcome was the setting up of a Curriculum Steering Committee to recommend an overall curriculum policy to the whole school community. The need for this committee arose directly out of concerns about curriculum planning raised by PARB. The CSC is at present examining

*Continued on page 25*

# Reviews



## Racist ideology

**IQ, Heritability and Racism: A Marxist Critique of Jensenism** by James M. Lawler. Lawrence and Wishart (1978) pp. 192

There is, perhaps, no more important question than that of human intellectual capacity, and especially the degree to which it is capable of change and development.' So begins Brian Simon's introduction to this excellent analytical study by the Professor of Philosophy at the State University of New York. At a time when the heredity thesis of human development is once again being promoted by right-wing and racist forces in both the United States and Britain, it is good to have a book which demolishes that thesis with such clarity and precision.

A little over a decade ago, in the early nineteen sixties, we may have been forgiven for thinking that the battle was already won here in Britain. The idea promoted by the late Sir Cyril Burt and others that intellectual capacity was wholly due to genetic endowment seemed to be discredited. There was wide-spread opposition to intelligence testing; and, at the same time, more and more teachers and educationists accepted the need to abolish streaming in primary schools and reorganize secondary education along comprehensive lines. An optimistic view of human potential even found its way into government reports concerned with education: in the words, for example, of the Newsom Report (1963), 'intellectual talent is not a fixed quantity with which we have to work but a variable that can be modified by social policy and educational approaches . . . the kind of intelligence which is measured by the tests so far applied is largely an acquired characteristic'. A thesis offering nothing but fatalism and despair had been rejected by many in favour of the reinstatement of *education* in the centre of the picture.

But the forces of reaction were not to be easily cowed. Early in 1969, Arthur Jensen wrote an article in the *Harvard Educational Review* entitled 'How Much Can We Boost IQ and Scholastic Achievement?' The title may seem innocuous enough; the conclusions drawn were quite venomous. It wasn't simply that Jensen was restating the 'classic' view of intelligence; more than this, his article was giving racist theories of human evolution and a spurious appearance of scientific respectability.

The position can be stated quite baldly: Jensen and his followers believe in the innate intellectual inferiority of working class and especially of black children. The starting-point of their argument is that there is a consistent finding of racial and class differences on IQ test; this thesis is then used to demonstrate the futility of massive compensatory education programmes designed to equalize opportunities and improve the average performance of the under-

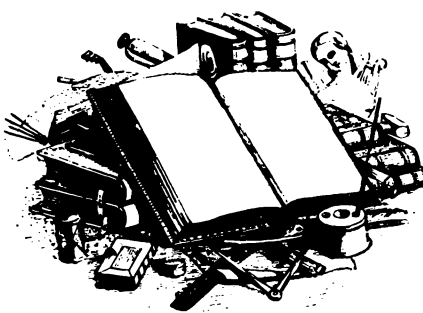
privileged. What, after all, is the point of educating black children on a large scale if the level of 'intelligence' that they reach, or fail to reach, has been basically decided once and for all in the genes?

In his Marxist critique, Professor Lawler deals comprehensively with the theory and practice of testing. He questions the reliability and validity of IQ tests; and rejects the assertion that 'thinking' is something that cannot be taught. He understands the history of mental testing, and shows how 'Social Darwinism' — the application of biological laws to the explanation of human evolution — continues to underlie the 'meritocracy' theory of Jensen and his followers. To accept the Jensen thesis is, after all, to believe that the few on top in society are there because of nature, because they were born with the brains; while the rest, either in the middle or at the bottom, have only themselves, or their genes, to blame.

Above all, Professor Lawler emphasizes the crude social and economic implications of Jensen's arguments. Referring to the American scene, he writes: 'in a situation where there is scarcity not of *abilities* but of *places* in higher education for youth without substantial economic means, the insinuation that special programmes or priorities that still exist for minority youth are being wasted on the intellectually inferior promotes racist division between blacks and whites. The concept of the intellectual inferiority of blacks encourages whites to fight blacks for shrinking educational opportunities, rather than to defend the legitimate special demands of blacks and to oppose the vicious system of racism.' This, surely, is the most frightening aspect of Jensen's work. Ideas are not matters of purely theoretical interest; they can have a profound effect on the way people treat one another. What Jensen has done is to contribute massively to the ideology and practice of racism.

CLYDE CHITTY

*Earl Shilton Community College, Leics.*



## Evaluation guide

**Evaluation and the Teacher's Role**, ed by Wynne Harlen, School Council Research Studies, Macmillan Education (1978) pp. 189 £6.95.

This book was written primarily for teachers who are concerned with evaluating in order to 'assist taking decisions about the organization, methods and content of work in schools'. It is, also of interest and value to others who may be involved with evaluation within schools; LEA advisory staff, college lecturers or anyone dealing with pre- or in-service training of teachers. The introduction makes clear the intentions and purposes of the book. It is seen as a resource book which can be used to give ideas, information and practical guidance on evaluation procedures which relate to the particular needs of those people concerned with taking decisions at whatever level within the school context. The short summaries of each chapter in the introduction are particularly helpful to teachers and others who may only be concerned with evaluation and decision-making for one aspect of school life at any particular time. These summaries direct the reader to the relevant contribution(s) so that sections of this book which may be of interest but do not relate directly to the problem in hand can be left till later.

The message which is reiterated throughout this book is the need for teachers to evaluate in order to make *decisions*. The collection of information without action on the basis of that information is considered a relatively useless operation. The contributors clearly see the teacher playing a vital role both in collecting information and in making the decisions concerning the organization, methods and content of work in the school. In order to make decisions information is needed and this book sets out ways of collecting and recording information, and the possible uses of this data in decision-making. The types of information described and their methods of collection may surprise readers who are used to thinking of data for evaluation in terms of conventional tests and examinations. Methods described range from the technique of 'focused observation' of individual pupils for assisting classroom-based, team-teaching decisions to the idea of indices for areas such as 'school prestige'; for example, 'do more pupils apply for admission than there are places available'.

The first two chapters, 'Evaluation and individual pupils' and 'Recording the progress of individuals' discuss evaluation in the context of making daily classroom decisions about pupils' learning experiences and methods for maintaining an efficient and effective recording system. In each chapter the importance of 'evaluation' is discussed, examples of various techniques and proce-



dures are given and a section on the questions which are raised by the preceding discussions are given for the reader to consider. The chapter dealing with 'Classroom accountability and the self-monitoring teacher' has a different emphasis. This discusses the concept of accountability and arising from this discussion is the notion of self-evaluation by teachers. This is illustrated from the work of the Ford Teaching Project. In this action-research project teachers were helped to clarify the consequences of their actions in embracing certain aims through the procedure of triangulation, where records and accounts of the same events from the teacher, pupils, and an observer are compared in order to help the teachers become aware of any gaps or inconsistencies of behaviour when compared with their intentions.

The shift from traditional curriculum development of producing materials towards the training and retraining of teachers and changes in organization which are required if new ideas rather than new materials are to be implemented are discussed in 'Organization for learning'. General strategies are described, and these are related to four specific examples: curriculum integration, mixed-ability grouping, team-teaching and open-plan organization. Different kinds of questions are suggested for each of these, but the authors make clear that these are dependent on the particular context and value system operating so it is impossible to draw up a blue-print for evaluation. The next contribution deals with some of the factors which affect 'curriculum decisions' but may not be acknowledged: the values held by those who take decisions, the information available and the constraints both human and material which operate within the context where decisions are taken. Sources of written information relevant to curriculum decisions are suggested which might not be immediately considered, as well as a very practical discussion of the constraints which operate, the need to recognize priorities for allocating time, materials, money and other resources in making curriculum decisions.

'The evaluation of the school as a whole' gives an overview of the areas of decision which need evaluative information and of methods which have been used. The importance of fitting evaluation to a purpose is discussed and the chapter then discusses what a school can do for itself. It includes a very comprehensive list of possible items for investigation which will help a school focus on a particular area for evaluation and decision-making. The final contribution summarizes issues which have been raised, re-emphasizing that evaluation and decision-making are part of the teacher's role, and considering questions such as accountability and communication in relation to evaluation.

This is not a prescriptive book, the responsibility for choosing the appropriate area for evaluation, the techniques and procedures for collecting and recording information,

and the nature of decisions which are taken are firmly placed with the teacher(s). It is however, practical, providing useful references which give more detailed information of the examples described should teachers wish to begin some kind of evaluation within their own schools. It is clearly argued too that if teachers are to be held accountable for what happens in classrooms and schools then they must be given the tools and expertise to undertake the evaluation and participate in the decision making which will result from that evaluation.

ANNE JASMAN  
*University of Leicester*

## Basis for action

**A School-Based Evaluation:** Documents from a DES Regional Course.

Cambridge Institute of Education Classroom Action Research Network, *Bulletin* No 3 (Spring 1979) 120pp

**In-School Evaluation** by Marten Shipman Heinemann Organization in Schools series (1979) 187pp £5.95.

**School-based Evaluation** is essentially a conference report. About one third of its length is concerned with membership details, the remainder with conference papers, documents arising from small group work and case studies. It suffers, as do all conference reports, from being rather fragmented. No doubt conference members would have a greater sense of cohesion and flow.

The conference papers, Pupil Assessment by Wynne Harlen, Methods for the Self-Analysis of Teaching (The Self-Assessment of Teacher Performance) by John Elliott and suggestions for a School Self-Evaluation Based on Democratic Principles by Helen Simons is the most valuable section. Each reflect the economic style of the lecture with plenty of highlighting of key points and the inclusion of diagrams and tables. What follows is rather less valuable and it is necessary to dig out the interesting bits. It must have been a lively conference and members will obviously find the report a useful record. A general reader will find parts of the report stimulating and useful.

Marten Shipman says of **In-School Evaluation** that 'it is a do-it-yourself manual.' He has painted a very broad canvas indeed and as a practising head teacher I have rarely felt so positively that my work will be enhanced if only I can follow a few of the leads he has given.

Evaluation is much talked about. Quoting the author out of context 'It is an area where rhetoric is not matched by action'. This book provides a basis for action. Not by providing exact models, but by suggestions, hints, ideas and examples. A very strong case is made for in-school evaluation and even the most timid will feel more confident.

I hope the author will not object if I momentarily compare his book with a 'good' cook book that I sometimes use on one of my very rare excursions into the kitchen to prepare a special meal. The guidelines are followed and with a little advice from the

local 'expert' an amazingly good dish can result. So many of us, even in the professional context of our lives, sometimes lack confidence. We know what we ought to be doing, even have a general idea how to set about it, but somehow don't get started. No reader of this book who wishes to establish or improve techniques of evaluation need worry about getting started.

Marten Shipman opens with a consideration of evaluation as 'a basic management tool' serving 'as an aid to decision-making.' Three chapters follow on school assessment. A number of useful books on assessment already exist so little is added to the store of assessment and testing techniques although a comprehensive and useful summary is given. The third of these chapters, however, deals thoroughly with that treacherous area concerned with the presentation of public examination results. Practical ways of presenting and using examination data are developed.

Although the first part of this book is so useful and interesting I found the final chapters on evaluating the wider aspects of school life, school organisation and curriculum, evaluation in context and the organization of in-school evaluation very stimulating. Much of this is new ground. Ideas and examples abound. Ways of collecting and presenting data are suggested. We are left in no doubt of the value of collecting detailed statistics, of presenting them clearly and of finding ways of evaluating all aspects of school life. 'The crucial part is the organization of an information base for the school so that staff can get a picture of trends in the school and comparison with external data.'

This is an extremely valuable, practical book that must be included on the essential reading list for those concerned with management in schools.

ROGER SECKINGTON  
*Earl Shilton Community College, Leics.*

### A Subscription to Forum

To: Forum,  
11 Beacon Street, Lichfield.

- \* Please send three issues of **Forum**, starting with Vol \_\_\_\_\_ No \_\_\_\_\_ to the address below.

I enclose cheque/P.O. for £2.50.

- \* Please send a Banker's Order Form to the address below.

NAME \_\_\_\_\_

ADDRESS \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

**The following Back Numbers of  
Forum are still available  
price 85p each**

<b>Vol 8 No 2</b>	<b>Special number on the Probationary Year.</b>
<b>Vol 9 No 3</b>	<b>Plowden symposium.</b>
<b>Vol 10 No 1</b>	<b>The Sixth Form in the Comprehensive School.</b>
<b>Vol 11 No 2</b>	<b>Two Years after Plowden; Self-directed learning.</b>
<b>Vol 11 No 3</b>	<b>Freedom of Choice – for whom? A new Education Act</b>
<b>Vol 12 No 2</b>	<b>From secondary to primary.</b>
<b>Vol 12 No 3</b>	<b>Teaching Unstreamed Classes.</b>
<b>Vol 13 No 1</b>	<b>Teachers for Comprehensives; Mixed ability science.</b>
<b>Vol 13 No 2</b>	<b>Assessment – for Whom? Recent trends in examining.</b>
<b>Vol 14 No 2</b>	<b>Innovation in Education.</b>
<b>Vol 15 No 1</b>	<b>Democracy and Innovation in Education.</b>
<b>Vol 15 No 2</b>	<b>16 to 19.</b>
<b>Vol 16 No 1</b>	<b>Resource-based learning.</b>
<b>Vol 16 No 3</b>	<b>Going Comprehensive in England, Wales and Scotland.</b>
<b>Vol 17 No 2</b>	<b>New Directions: reconstruction of knowledge.</b>
<b>Vol 17 No 3</b>	<b>The Question of Size for primary and secondary schools.</b>
<b>Vol 18 No 1</b>	<b>Mixed Ability Teaching: French, Maths, Science.</b>
<b>Vol 18 No 2</b>	<b>Flexibility and Mixed Ability.</b>
<b>Vol 18 No 3</b>	<b>Examination or Assessment in primary and secondary schools.</b>
<b>Vol 19 No 1</b>	<b>In Defence of Education.</b>
<b>Vol 19 No 2</b>	<b>Comprehensive Remedial Provision for primary and secondary.</b>
<b>Vol 19 No 3</b>	<b>The Primary School.</b>
<b>Vol 20 No 1</b>	<b>Multiracial Education.</b>
<b>Vol 20 No 2</b>	<b>Non-Streaming – Why and how.</b>
<b>Vol 20 No 3</b>	<b>Testing and Teaching.</b>
<b>Vol 21 No 1</b>	<b>New Opportunities: the lower birthrate.</b>
<b>Vol 21 No 2</b>	<b>Primary and Secondary.</b>
<b>Vol 21 No 3</b>	<b>Mixed Ability Teaching and Learning.</b>

**Order from            Forum, 11 Beacon Street, Lichfield WS13 7AA.**