

42

Hence the importance of holding fast to true educational values, and not yielding to false arguments. For it is simply not true that a vocationalised curriculum either secures jobs, or increases economic wealth. A report for the Department of Environment¹² has found that vocational preparation and training have 'only a marginal effect on levels of employment'. Japan — a country much admired by the technocrats — has secondary schools with 'no vocational courses for 14-year-olds, no work experience for 15-year-olds . . . Technology teaching is primitive . . . there are no computers in the classroom . . . All children follow a broad, general course.'¹³ And it is worth noting that in Sweden, 'direct vocational training cannot by law be provided in the pre-16 comprehensive schools.'¹⁴

If the dire toxin of vocationalism spreads through our secondary schools, it will bring in its wake only social divisiveness and deficient forms of schooling. But it can only spread if teachers succumb to its seductive but meretricious message. Schools should do all in their power to resist the fractured logic of the 14-18 curriculum, to strengthen the boundary at 16-plus, and to devise school-based 11-16 core curriculums which build not on differentiation but on unity. Nothing less will equip their pupils for the world which lies ahead.

References

- 1 W A Reid (1976), letter to *The Times Educational Supplement*, 24 October.
- 2 Grubb, W and Lazerson, M (1981), 'Vocational solutions to youth problems: the persistent frustrations of the American experience', *Educational Analysis*, summer.
- 3 'The Youth Training Scheme' (1982), *Bulletin*, Youthaid, December.
- 4 *A New Training Initiative* (1981), Cmnd 8455, HMSO.
- 5 Macintosh, H (1983), 'Down with 16-plus', *The Times Educational Supplement*, 11 February.
- 6 Described in Holt, M (1981), *Evaluating the Evaluators*, Hodder and Stoughton.
- 7 Her Majesty's Inspectorate (1983), *Report on Graded Tests in Modern Languages in Oxfordshire*, HMSO.
- 8 Scottish Vocational Preparation Unit (1982), *Assessment in Youth Training: made-to-measure?* Jordanhill College of Education, Glasgow.
- 9 Peters, R S (1967) *The Concept of Education*, Routledge.
- 10 HMI (1979), *Aspects of Secondary Education*, HMSO.
- 11 Reported in *The Times Educational Supplement* (1983), 4 February.
- 12 Markhall, G and Finn, D (1982), *Young People and the Labour Market*, William Temple Foundation for Department of Environment.
- 13 Stevens, A (1982) 'Lessons Japan can teach the West', *Observer*, 19 December.
- 14 Owens, E (1981) '16 to 19 in Sweden', *Education*, 27 February.

The Teacher's View of Testing

Forum, Vol.25, No.3, Summer 1983

Stephen Steadman and Harvey Goldstein

Stephen Steadman and Harvey Goldstein report here on material derived from the Evaluation of Testing in Schools project, funded by the Social Science Research Council at the Institute of Education, London University.

'There is no standard of comparison which can surpass or supercede the considered estimate of an observant teacher, working daily with the individual children over a period of several months or years. This is the criterion I have used'. (Burt, C. *Mental and Scholastic Tests*, LCC, 1921, p.199)

' . . . when all the teachers were considered together, almost half of them (forty eight per cent) were inconsistent estimators, who over-estimated about half their pupils' performances, while under-estimating the rest. Of the remaining fifty two per cent of the teachers, about two-thirds were under-estimators, the rest being over-estimators'.

(Southgate, V., Arnold, H. & Johnson, S. (1981) *Extending Beginning Reading*, Heinemann Educational Books, p.93)

We can no longer be sure whether Cyril Burt actually did what he said, but the contrast between his expressed attitude and the attitude behind the work of the Schools Council team — which contrasted teacher estimates of reading age with those derived from Schonell's Graded Word Reading Test — is dramatic and unmistakable. For Burt, test results should give way in the face of teacher estimates; for Southgate and her team, 60 years later, the teachers' estimates were judged less accurate than a test result.¹ This article describes our findings on the question of how far teachers are prepared to believe test results.

The Evaluation of Testing in Schools Project (ETSP) was funded by the SSRC for three years from January 1980 as a part of the SSRC's programme of research into aspects of accountability in education. The project's general aim has been to determine the extent of standardised testing in schools, the reasons for introducing testing, the uses to which test results are put, and the effects of testing on schools in the broad sense. Two particular foci have been an evaluation of the work of the DES Assessment of Performance Unit² and LEA test programmes which typically set out to 'screen', 'monitor' and 'aid transfer' as well as having accountability purposes.

Early questionnaire surveys and visits to LEAs showed that almost eighty per cent of all LEAs test at least one age group, mostly using 'blanket' or saturation testing.³ Reading is the most commonly tested skill and there is evidence that, since the recommendations of the Bullock Report, newer and better tests are being used by LEAs.⁴

More recently we have conducted an interview survey in a random sample of 20 LEAs to obtain heads' and teachers' views about testing at local and national level. The sample of LEAs was drawn from the seventy eight LEAs known to have testing programmes using standardised tests. In each LEA a random sample of four schools with junior age pupils was visited. All the

LEAs tested at junior age levels; a minority also tested at the secondary school level. Within each school we sought interviews with the head and two teachers. In principle the teachers were those who taught the age group(s) at which the LEAs did their testing. (LEA testing is most often done in the first and last years of junior schooling.) But, in practice, this was not always possible because of differences in school organisation and size. The sample produced interviews with eighty heads and 158 teachers. This article presents some preliminary results from an analysis of these interviews.

Teachers who disagree with test results

The heads and classroom teachers were asked about their attendance on initial or in-service training courses on educational testing, since we felt that knowledge about the nature of testing would help teachers make a more rational use of test results. In the event about two-thirds of the heads and class teachers had been on a course at least part of which was devoted to testing, and about a third of the heads and a fifth of the class teachers had been on a whole course specifically devoted to testing.

We were interested in teachers' trust in test results. They were asked firstly whether they ever found discrepancies between their own ratings of children's attainments and the results of a test. Over ninety per cent said that this did occur. We then went on to examine what teachers did when this occurred — whether they tended to believe the test or their own judgement. The same questions were asked of both heads and class teachers. 'If the test score is higher than you expect, what is your reaction?'. An equivalent wording was then used to ask about reactions if the score was lower than expected. Answers ranged quite widely in the amount of detail given, with the headteachers being generally more fluent, and mentioning more possibilities in their replies than did the class teachers. Also the headteachers tended to assume that verbal reasoning tests were the subject of discussion, whereas teachers more often referred to reading tests. This may reflect who does what in primary school testing. One head teacher said that, if the test score was higher than he'd expected, his reaction would be:

'That he's a lazy little devil! It would be rather a blow to my professional judgement . . . I'd be nonplussed really, if the child scored *very* high and I was taken by surprise. I don't regard these scores as sacrosanct though. I think, within five or ten points, you can *expect* little variations'.

And if the score was lower than expected:

'I tend to be sympathetic. I'd think — and say, well, perhaps he wasn't feeling very well that day. I try not to place too much weight on a low score'.

Another head had a less relaxed view. If the score was unexpectedly high he'd:

' . . . believe the child is underachieving and find out why — (I'd) retest. Maybe call in the educational psychologist or remedial (advice). . . . see (the) home and check there. Talk to (the) parents'.

And, if the score was unexpectedly low, he'd:

'Check health and nervousness. Children can overachieve, I suppose. Talk, and find a consensus about the child. Retest with another test to check with other teachers'.

The interview schedules carried pre-coded categories of answer and it later proved possible to add further categories by inspecting the answers which the interviewers had been instructed to note down verbatim. We also used the open ended replies to form a judgement as to whether the discrepant scores were believed or not, although sometimes a clear-cut decision was not always possible.

If we confine our interest to those 48 heads and 106 class teachers where a clear decision was possible, and compare reactions when the score is higher than expected with reactions to lower than expected scores, an interesting result emerges. When the test score is higher than a teacher's own expectation, about three-quarters of the heads and half the teachers believe the test score to be correct and when the test score is lower than expected the same percentages believe it. However, among the heads, nearly all maintain a consistent attitude, either believing the test both when higher and lower than expectation or refusing to believe it whatever its result. Some class teachers on the other hand have an apparently inconsistent attitude. About one-sixth believe the test score when lower but not when higher than expected, and a similar percentage believe it when lower, but not when higher, than expected.

Thus we have a picture of heads more ready to trust tests than teachers, perhaps a reflection of their relative distance from the classroom situation. In addition they are consistent in their views. Class teachers on the other hand exhibit more doubts. Those who believe the test score when higher than expected, but not when lower, could be said to be exercising their professional judgement to give a child the 'benefit of the doubt' in the realisation that it is better to have a routine anticipation of higher achievement even if 'in error'. Those who believe test scores when lower than expected, but not when higher, could well be those whose expectation is based on what they know is the child's best achievement, rather than his average performance. Such teachers would thus view with suspicion a test score higher than such a high expectation but would have no difficulty with a lower than expected score.

In order to try to probe these attitudes further we studied teachers in terms of the amount of exposure to courses on testing and their teaching experience.

When broken down by whether the teachers had been on any course or not, substantially the same picture emerged, but when classified in terms of years of teaching experience some interesting differences appeared. For the heads, classified into those with 0-20 years experience and those with twenty one years or more, those with less experience tended to show more 'inconsistency'. About one-fifth especially tended to believe higher than expected scores and disbelieve lower than expected scores. Among the longer experienced, there were no inconsistent heads. For the classroom teachers, classified into those with ten or less years experience and those with eleven or more, the opposite was the case. For those with less experience, there was an overall higher tendency to believe the test with only one-sixth showing 'inconsistency'. For the more experienced, only about half believed the test, the remainder being 'inconsistent'.

Conclusions

We are continuing to analyse these results (Gipps *et al*, 1983),⁵ but already some preliminary conclusions are in

order. Teachers' responses to a child's standardised test score are not simple. There are teachers who appear to accept these scores, even when contrary to their own judgement, and we have suggested that this may be related to the teachers' image of a child and how he or she frames their expectation. Head teachers seem more inclined to believe test scores than class teachers, especially those heads with longer experience. On the other hand, class teachers with longer experience are inclined to be more sceptical about the test results. Within our data there is no simple way to provide an explanation of these findings, and the following commentary makes only tentative suggestions.

There is other research from the United States which tends to corroborate some aspects of our findings. Very similar investigations to our own have been conducted by the Universities of Pittsburgh and Carnegie-Mellon. The work has included an interview survey of practising elementary school teachers and the most directly comparable findings have been summarised by Leslie Salmon-Cox⁶ as follows:

'When a test score indicates performance below that which a teacher would predict from classroom performance, the score tends to be discounted. When a child scores *higher* than might be predicted, it seems to serve as a "red flag" indicating that the teacher has missed something'.

So far as we are aware, this 'inconsistency' has not been investigated further by the Pittsburgh group. In commenting upon these findings, Cox says that standardised tests fall short, as far as teachers are concerned, in two ways. They only measure certain aspects of teachers' cognitive goals — almost nothing of the social goals which teachers rate highly — and they are not the broad-based kinds of measures that teachers prefer. George Madaus⁷ has remarked upon the anomalous position of standardised tests.

'If the results of the tests differ greatly from teachers' perceptions, the tests run the risk of being ignored on the grounds of inaccuracy. If, on the other hand, test results correspond closely to teachers' perceptions, the tests run the risk of being dismissed on the grounds of redundancy'.

Compared to our results, it would seem that fewer American teachers believe a test score which is lower than their expectations. It is possible that this is related to the increased exposure to testing and hence sophistication about tests of US teachers, although we are unclear as to how this would operate. If our earlier suggestion about expectations is correct, it would suggest that upgrades tend to be formulated differently in the two counties.

There are a number of ways in which discrepancies between test score and teacher expectation can arise. Most obviously, the test may be examining aspects of performance only loosely related to the curriculum in operation. Even where the test is relevant, however, the random 'measurement error' in some cases will be large enough to indicate a difference from expectation. Some heads and class teachers were aware of this effect, referring to it as 'a freak result', 'an element of luck' or 'a fluke'. More importantly than either of these two explanations perhaps is the likelihood that the teacher will be judging the child by local criteria — both in terms of curriculum and by comparison with other local children, whereas most test norms are national. In addition, the mismatch between these aspects will vary

according to the test used.

We see therefore, that there will be many occasions where the test and teacher expectation will legitimately differ and there was recognition of this by some of the teachers in our study.

To ask which is the 'right' assessment is to ask an irrelevant question since the two types of assessment have different aims. Nevertheless, we suspect that this is often not the description of testing which gets emphasised in courses and in some quarters there is often an assumption that a test is an 'objective' standard against which to measure the teacher. Our own view is that while tests have a part to play in assessment, both for 'monitoring' and 'screening', their role is not widely understood and nor is their actual mode of use. The present study has, we hope, indicated some of the extent of this deficiency.

References

1. See Stierer, B. (1982). 'Testing Teachers? A Critical Look at the Schools Council Project "Extending Beginning Reading"', in *Primary Education Review*, NUT, spring, No.13, pp.17-20. Barry Stierer is an SSRC Student 'linked' to the ETSP.
2. Gipps, C. and Goldstein, H. (1983). *Monitoring Children: An Evaluation of the Assessment of Performance Unit*, Heinemann, London.
3. Wood, R. and Gipps, C. (1982). 'An Enquiry into the use of Test Results for Accountability Purposes' in *Calling Education to Account*, McCormick, R. (ed), Heinemann Educational Books in association with the Open University Press.
4. Gipps, C. and Wood, R. (1981). 'The Testing of Reading in LEAs: the Bullock Report seven years on', *Educational Studies*, 7.2, pp.133-143.
5. Gipps, C., Steadman, S.D. and Blackstone, T. (1983). *Testing Children*, to be published by Heinemann.
6. Salmon-Cox, L. (1981). 'Teachers and Standardised Achievement Tests: What's Really Happening?' *Phi Delta Kappan*, May, pp.631-133.
7. *Phi Delta Kappan*, May 1981, contains a special section on standardised testing.

OBITUARY

RAYMOND KING

The death of the chairman of our Editorial Board, Raymond King, will be felt as a sad loss by several generations in the mainstream of progressive education. Long associated with the English New Education Fellowship and *New Era*, Raymond was a founder member of *Forum's* editorial board, contributing a major article, 'The London School Plan: the present stage', to the first number of *Forum* 25 years ago. He became chairman of the board in 1964, guiding discussion with wisdom and humour for nearly 20 years.

Raymond was an influential leader in developing universal secondary education. Appointed a grammar school head in 1926, he was part of a caucus of London heads planning a vision for the future in the 1930s, chairman of the standing Conference on the Democratic Reconstruction of Education in the next decade and wrote the ENEF pamphlet, *The Comprehensive School*, in 1950. Six years later he began turning Wandsworth School into a full comprehensive. In his last major *Forum* article (vol 22 no 1) he surveyed five decades in the evolution of comprehensive education from his own central standpoint.
Ed.