

TESTING WITH EDUCATIONALLY DISADVANTAGED CHILDREN

A.E.G. Pilliner,
Director, Godfrey Thomson Unit for Academic Assessment,
University of Edinburgh, Scotland.

Every country has its educationally disadvantaged children, even those in which educational development is most advanced. Britain is now replacing a selective system of education by a comprehensive one in an endeavour to eliminate, or at least reduce, unequal educational opportunity. Superimposed on this task, she is now faced with the responsibility of educating an increasing number of young immigrants from other Commonwealth countries. The United States of America, despite the fact that in principle her educational system has never been other than comprehensive, has not yet achieved her avowed aim of de-segregation and the quality of education offered to some of her citizens is still inferior to that enjoyed by others.

It should not surprise us that this state of affairs exists a fortiori in countries at an earlier stage of development. In such countries, through sheer force of circumstances, education in any way comparable in quality and adequacy to that taken for granted for the majority in some advanced countries is available to only a small minority. The allotment of a large proportion of scarce resources to the more extensive education of a relatively small proportion of children is understandable. Pre-requisite to speedier progress in the future, further technical advance, increased economic development and wider educational expansion, is the production now of a necessarily small number of people possessing the knowledge, skills and dedication essential to the achievement of these aims.

It is no accident that in many of the new countries educational objectives tend increasingly to resemble those in others more fortunate in having advanced further along the path of development. According to Doob (6), the pressures forcing the new countries in the same direction are inevitable, irresistible and irreversible. This does not mean that all will arrive at the same place. A country on its way 'up' will be selective in what it absorbs and will adapt its acquisitions from elsewhere to its own traditions and needs. Nevertheless, since both less and more developed countries share a number of the same objectives, their educative processes will have much in common. At the same time, a process evolved over a lengthy period and geared to the norms of a society or culture already well developed cannot be transferred ready-made to another less so without giving rise to problems, even though the objectives are similar.

These problems are reflected in the testing procedures which are an integral part of any and every educational process. The situation previously mentioned implies that a relatively small number of pupils must be selected for secondary education from a very large primary school population. Countries where this situation exists are likely to be characterised by primary education of poor and uneven quality. Children living in towns may be more fortunate in their primary education than others living in villages.

In this case, restrictions are imposed on the interpretation of scholastic attainment test results. Though such tests may still accurately

measure a pupil's achievement in specific subjects to date, their use as prognosticators of future success is precluded or at least limited, however successful they may be in this respect with children more fortunately circumstanced. The poor performance on an arithmetic test of a pupil who has hitherto been taught arithmetic either badly or not at all is a fair index of his present ability in that subject. But as a predictor of his likely progress if this defect is remedied its value is questionable. If assessment of potential or aptitude is at issue some other means must be found.

If experience in countries more educationally advanced is anything to go by, the use of tests of verbal reasoning might seem to offer a solution. Tests in this category differ from tests of scholastic attainment in that they are less closely geared to the school curriculum; good performance on them is less dependent on exposure to the usual range of school subjects. They have been extensively used for 11+ selection in Britain, where numerous follow-up studies have consistently shown them to be among the best predictors of academic success.

However, difficulties still remain. There may be several native languages or dialects while the accepted medium of instruction in the secondary schools is a second language such as English, the pupil's acquaintance with which is limited by factors such as his primary teachers' command of it.

Bernstein's (3) work on language habits in Britain bears on this situation. He points to the relation between class structure and the varieties of English used by school children. Social stratification is related to differential availability of language codes. The lower working class child has a group-oriented 'restricted' code; the middle class child has both this and an individually-oriented 'elaborated' code. These codes differ in that the first is more fluent, repetitive and predictable, the second more hesitant, idiosyncratically planned and complex. Educationally, the child from a poor background is at a disadvantage since he finds himself having in effect to translate what he hears from his teachers. As Bernstein points out, differential difficulty in communication is likely to be reflected in differential verbal test performance.

The problem is exacerbated when the differential is not merely intra- but inter-language. It is therefore natural to consider the possibility of assessing pupils' aptitude for further stages of education by some testing procedure which avoids the use of the differentially unfamiliar second language. On the face of it, one way of doing so would be to couch the tests employed in the pupils' own native languages. This however may be difficult in practice if several languages are involved. There is the further technical difficulty of equating the performances of different children on different tests - for, let there be no mistake about it, even the same content translated into different languages produces different tests, the results of which, expressed numerically, are not necessarily comparable. Moreover, the problem of unequal primary school opportunity, and its implications for scholastic attainment, will still remain.

On all these counts, it may be thought desirable to go one step further, to eliminate the use of language so far as is practically possible, and to rely on non-verbal or non-language tests. Here, surely, it might be argued, is the way out of the difficulty. If the use of language-bound tests is seen as impracticable or leading to injustice, should not their substitution by non-language tests reduce the practical problems and promote

'fairness' for all concerned?

This is the kind of thinking behind the more general concepts of 'culture-free' and 'culture-fair' testing. The intention is wholly admirable. Any measure which will help to redress the balance in favour of children who are culturally deprived or otherwise educationally disadvantaged is surely to be encouraged. The laudable objective is to reduce these obstacles by the use of testing devices which transcend or remove cultural differences or educational inequalities.

However, the problem is by no means simple. The concept of 'culture-free' tests is highly dubious. Anastasi (1, p.256) is surely right when she says: 'No test can be truly "culture-free". Since every test measures a sample of behaviour, it will reflect factors that influence behaviour. Persons do not react in a cultural vacuum.' Wesmen (16, p.269) is even more forthright. 'I do not wish to impugn the high social motives which stimulate the search for such devices; I do wish to question that such a search, in its usual setting, is sensible. A culture-free test would presumably probe learnings which had not been affected by environment; this is sheer nonsense.' These statements represent the general view of most contemporary psychologists. Few would now regard the quest for culture-free tests as other than chimerical.

The prospect for 'culture-fair' tests is, on the face of it, less unpromising. In principle it is possible to build tests which, though not free of cultural influences, sample only behaviour common to several cultures. An alternative description of such tests is 'cross-cultural'. The amount of effort that has gone into the construction of allegedly cross-cultural tests is vast, particularly if we include also tests intended for comparisons among sub-cultures within a larger culture. Only a few can be mentioned here. In the nature of things, they are non-verbal in content. They fall into two main categories: performance tests, designed for individual administration, and in the main involving manipulation of objects; and non-verbal or non-language group tests, normally paper-and-pencil tests which do not demand of the testees the skills of reading and writing. Most such tests do however depend on oral instructions, it being assumed (perhaps too lightly) that these are of such simplicity that no semantic problems arise in their translation and that different language versions do not differ in difficulty. A few tests have been constructed in which the instructions can be mimed or demonstrated.

Examples of tests in the performance category are: Form-board (Sequin, Pintner-Paterson), Mazes (Porteous), Picture Completion (Healy), Block Manipulation (Kohs); Stencil Design (Arthur); Analogies (Leiter); and, of course, the General Performance Scale of the WISC (Wechsler). Examples from the group non-language category are the Draw-a-Man (Goodenough), Matrices (Raven), Pictorial Problems (Davies-Eells), Semantic Symbols (Rulon); and a number of tests intended to probe, using pictorial or diagrammatic material, mental functions - analogies, odd-man-out, series and the like - similar to those frequently occurring in verbal tests (Moray House Picture, Jenkins Non-Verbal, Cattell IPAT).

On closer examination, however, the prospect of producing 'culture-fair' tests is only slightly less unpromising than for tests that are 'culture-free'. By restricting test content to elements common to several cultures the relevance of the results in respect of any one of them is made questionable. To the extent that different cultures display unique features,

nurture disparate traditions and values, or foster or suppress different abilities or modes of behaviour, tests restricted in this way may miss their targets. To quote Anastasi (2, p.299) again: 'If we were to rule out cultural differentials from a test, we might thereby lower its validity against the criterion we are trying to predict'. It is as though in trying to please everybody, we succeed in pleasing nobody. Or, to change the metaphor, although the wave pattern for the fundamental tone emitted by different musical instruments is the same for all, it is the superimposed over-tones or harmonics which endow each with its peculiar timbre, its richness of quality.

The concepts of 'culture-free' and 'culture-fair' tests once received plausible support from the contemporary psychological theory. 'Native intelligence', like original sin, was reified and came to be regarded as a fixed entity rather than a developing attribute. By the exercise of sufficient inventiveness - Wesman (16) speaks of 'ingenious mining devices' - the influence of differential exposure to learning could be eliminated and the 'innate intelligence' of the individual revealed and recorded on a scale for all to see.

More recent theory is less accommodating. Hebb's (9) distinction between Intelligence A and B corresponds broadly to the geneticist's distinction between genotype and phenotype. Like the genotype, Intelligence A is not directly observable, still less measurable. Only Intelligence B, corresponding to the phenotype, can be observed; it results from the interaction of both nature and nurture. The title of a once popular song sums it up neatly; 'It's what you do with what you've got that counts'. Vernon (14) playfully, in the first place, one suspects, but then more seriously, had added a further category. Intelligence C is what tests measure. It varies with difference in test content and is therefore not unique in the prediction it affords of Intelligence B. Hebb's theory offers but cold comfort in the search for instruments equally fair to differentially disadvantaged testees.

On the fact of it at least, the theory of 'fluid' and 'crystallised' intelligence attributable to Cattell and Horn (5) is distinctly more hopeful. They suggest that the general factor emerging from studies of batteries of disparate tests is a mixture separable into two components: G_f ('fluid' intelligence), reflecting constitutional equipment; and G_c ('crystallised' intelligence), the results of experience such as cultural and educational pressures. Unlike Intelligence A, G_f is measurable by tests tapping adaptability to situations so unfamiliar that previous learning experience is of no help. G_c , corresponding roughly to Intelligence B, is manifested in cognitive behaviour already patterned by previous experience. Even before biological maturity is reached, diversity in cultural opportunities, interests and personality traits produces substantial individual differences in G_c which, according to the theory, should not be paralleled for G_f .

This theory underlies the construction of the Cattell IPAT Culture Fair (formerly Culture Free) Intelligence Test. Predictably, the greatest success in removing 'contamination' by cultural differences is claimed for subtests involving mazes, identification of similar drawings, picture classification and symbol copying. At best, however, the success achieved is only partial. In view of the IPAT, Tannenbaum (12, p.454) concludes that 'the goal of demonstrating equality among national and international subpopulations by some measures of general ability has not been reached

by this test.' He questions whether this is a goal worth pursuing. 'Even if it were possible to devise a test so antiseptic as to clean out inequality not only among subcultures but also among other groups showing differences in test intelligence, such as those classified by sex, age, geographic origin, body type, physical health, personality structure, and family unity - what kind of instrument would we have then? Since such a test must perforce be so thoroughly doctored as to omit tasks that reveal these group differences, or substitute others that show "no difference", what could it possibly measure? What could it predict?' Vernon's (15, p.25) conclusions are equally definite. 'The main weakness in his (Cattell's) theory is the claim that fluid ability tests are largely immune to cultural influences. The skills required for reasoning with these abstract materials would appear to be built up in just the same way as those involved in verbal reasoning; and the evidence ... demonstrates at least as great variation attributable to cultural differences'.

For a very complete and up-to-date survey of this evidence, reference should be made to Vernon (15). Only some of it can be cited here. As already stated, the IPAT was found to be only partially successful in ironing out cultural differences. Although in cultures similar to that in which the test was developed the same norms were approximately applicable, this was not so for cultures more dissimilar; for these, average performance was often much lower. Bernstein (4) reports smaller differences in performance on Raven's Matrices between middle and working class groups than on tests of verbal reasoning. But in other studies, particularly in African countries, test results were positively correlated with amount of education. The Goodenough Draw-a-Man (8) test has gone through several revisions. After extensive use with a number of different cultural and ethnic groups, its authors have abandoned their original optimistic view and in their more recent reports have concluded that a culture-fair test of whatever attribute 'is illusory'.

The Davis-Eeels Games (7) were specially designed for American use to be relatively independent of social class bias. But differential educational disadvantage was still reflected in differential performance on these tests no less than on more conventional intelligence tests which were in addition more predictively valid in respect of tested achievement and teachers' assessments.

One of the most interesting and definitive studies in this area is that conducted by Ortar (10). She administered both a Hebrew version of the WISC Verbal Scale and also the Performance Scale to upwards of 1000 Israeli children. These were divided into five groups with different cultural backgrounds ranging from recently arrived Oriental immigrants to an Israel-born 'high status' group (mainly of European parentage). After re-standardising both Scales for Israeli children, she found the 'cultural distances' between the groups to be larger on the Performance than on the Verbal Scale. In a similar study conducted with Scottish children Tsakalos (13) found differences in social status to be reflected in differential performance on the Jenkins non-verbal test no less than on Moray House tests of verbal reasoning and scholastic attainment.

The conclusion is inescapable that it is fruitless to search for testing instruments that will somehow transcend cultural differences and educational inequalities. What are the implications?

In the first place, it must be recognised that belief in the essential

equality of man receives little support from the considerable research in this area which it has stimulated. It remains an act of faith. This need not deter us from acting on that belief. A warrant from psychologists qua psychologists is not essential to the maintainance of a fundamental principle on which the advance of civilisation is predicted.

Secondly, it has to be accepted that educational disadvantage is endemic and that there is no simple counter to it by way of tests purporting to reveal intelligence, talent, potential, or whatever we may choose to call it, irrespective of differences in cultural, social or educational background. Such tests are of dubious value to a primary school teacher in Britain faced with an influx into her class of immigrant children without a word of English among them. There is no simple way of helping her to differentiate among them, or between them and their native-born peers, in terms of 'basic' intelligence. Her best practical policy still is to do all she can to make them feel welcome and to teach them English. Likewise, such tests offer no panacea to a developing country where, because of scarce resources, stringent selection is necessary and too many children are chasing too few places in the educational sun. The brutal truth must be faced that there are plenty of other children whose claim for preferment is no worse than that of the fortunate few selected. The solution to the problem is economic, not psychometric.

From an educational stand-point, the best hope of advance in general, and amelioration of educational disadvantage in particular, lies in the field of language-teaching. The mother-tongue may suffice if it provides for effective communication with other nationals and is suitable as a medium for advanced education. If not, a second language is necessary, taught, as Vernon points out, not peripherally, but as a central tool of comprehension and thought.

What then should be the role of the psychologist? There is no reason why it should change materially, though possibly a shift of emphasis is indicated. Any still engaged in the search for testing instruments equally 'fair' in different cultures should bear in mind the fruitless quest of the alchemists for the philosopher's stone; though they may console themselves by reflecting that (in a different sense from the original alchemists') the transmutation of metals has now been accomplished. There is a lesson here. That achievement was the outcome of 'pure' research not specifically aimed at transmutation, nor concerned with its consequences. So too with the psychologist. He should listen to Anastasi's (2, p.302) warning: 'It is not (the psychologists') role to provide ready-made solutions to insoluble problems. It might be salutary if testing gave less heed to the pull of practical needs and more to the thrust of behavioural sciences'.

But less heed is not the same as no heed at all. The psychologist, like the physicist, has responsibilities outside his laboratory. Despite all that has been said, he has yet much to give in the field of testing in the service of education. It is a truism that the best indicator of a child's learning potential is a test sampling previous learnings which are relevant to the criterion or criteria we wish to predict. For long enough this maxim has guided with reasonable success the construction of tests for educational purposes within western cultures. There is still room for further research of the kind that Schwarz (11) has engaged in, aimed at discovering the previous relevant learnings in cultures elsewhere in a stage of transition.

Let Vernon (15, p.229) have the last word. 'What is important is that in concentrating on abilities recognised by western cultures,

psychologists should not neglect special talents that might be more highly developed in other countries'. To extend a metaphor employed earlier, in seeking out these special talents we may be taking a small but useful step towards the assembly of a cross-cultural orchestra.

REFERENCES

1. Anastasi, A. (1964) Psychological Testing. (2nd Edition). New York: Macmillan.
2. Anastasi, A. (1967) 'Psychology, Psychologists, and Psychological Testing'. In American Psychologist, 22,4, pp. 297-306.
3. Bernstein, B.B. (1961) 'Social Class and Linguistic Development: A Theory of Social Learning'. In Education, Economy and Society (Halsey, A.H.) Glencoe: The Free Press, pp. 288-314.
4. Bernstein, B.B., and Young D. (1966) 'Some Aspects of the Relationships between Communication and Performance in Tests'. In Genetic and Environmental Factors in Human Ability (Meads, J.E. and Parkes, A.S.). Edinburgh: Oliver and Boyd, pp. 15-23.
5. Cattell, R.B. (1963) 'Theory of Fluid and Crystallised Intelligence: A Critical Experiment'. J. Educ. Psychol. 54 pp. 1-22.
6. Doob, L.W. (1960) Becoming More Civilised. New Haven: Yale University Press.
7. Eeels, K., Davies, A., Havighurst, R.J., Herrick, V.E., and Tyler, R.W., (1951) Intelligence and Cultural Differences. Chicago: University of Chicago Press.
8. Goodenough, F.L. and Harris D.B. (1950) 'Studies in the Psychology of Children's Drawings': II. 1928-1948. Psychol. Bull. 47, pp. 369-433.
9. Hebb, D.O. (1949) The Organisation of Behaviour. New York: Wiley.
10. Ortar, G. (1963) 'Is a Verbal Test Cross-Cultural?' In Scripta Hierosolymitana. Jerusalem: Magnes Press, The Hebrew University, pp. 219-35.
11. Schwarz, P.A. (1961) Aptitude Tests for Use in Developing Nations. Pittsburgh: American Institute for Research.
12. Tannenbaum, A.J. (1965) Review of IPAT Culture Fair Intelligence Test. In Mental Measurements Year Book (Ed. Buros, O.K.) New Jersey: The Gryphon Press, pp. 453-4.

13. Tsakalos, P. (1966) 'Is a Non-Verbal a Culture Fair Test?'
Master of Education Thesis (unpublished).
University of Edinburgh.
14. Vernon, P.E. (1960) Intelligence and Attainment Tests.
London: University of London Press.
15. Vernon, P.E. (1969) Intelligence and Cultural Environment.
London: Methuen.
16. Wesman, A.G. (1968) 'Intelligent Testing'. In American Psychologist, 23, 4, pp. 267-274.