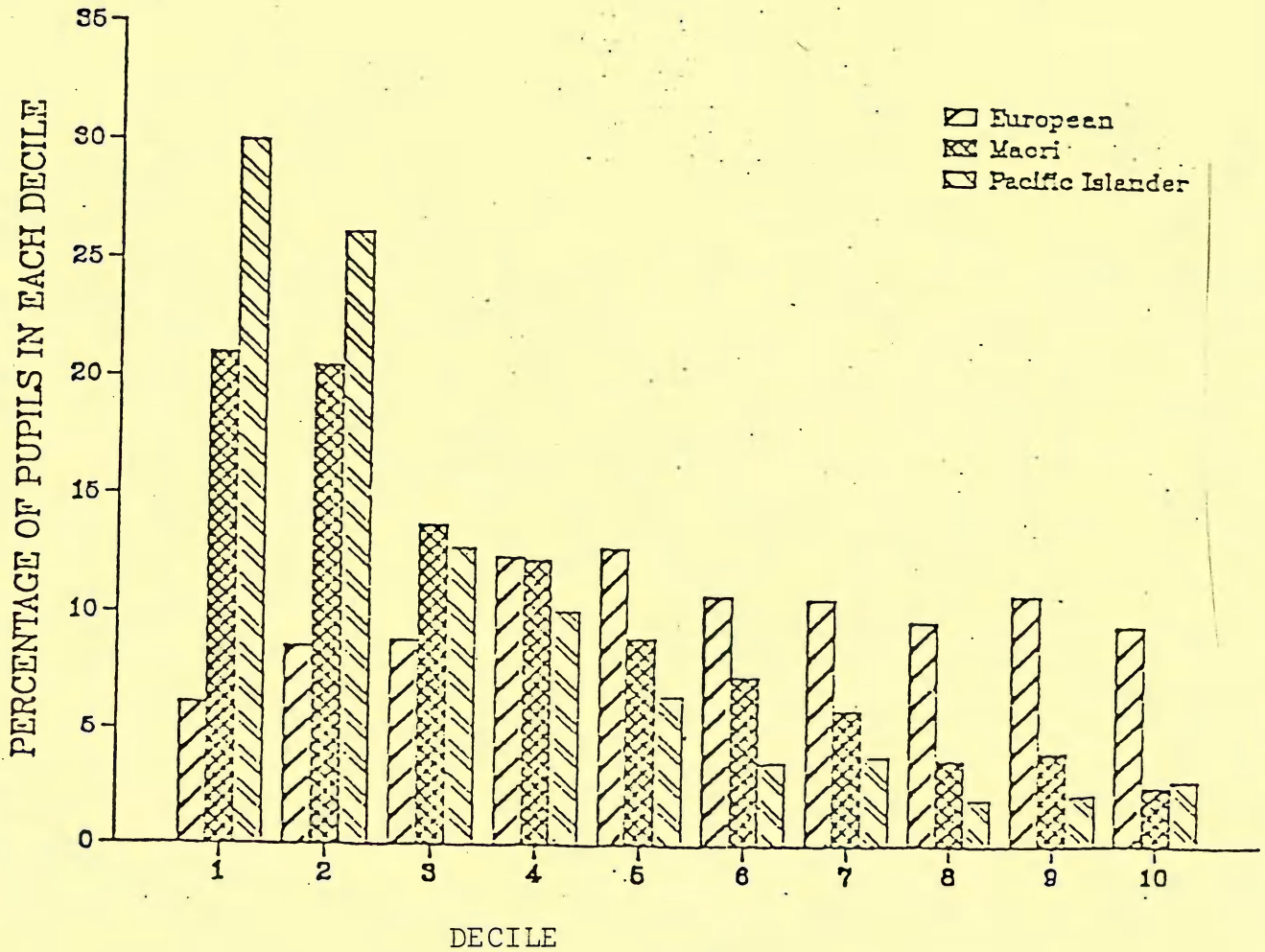


TESTING TESTING

TOSCA DECILE PERCENTAGES



ISSUES IN EDUCATION - number 12

Programme On Racism,
NCC Box 9573, Auckland.

from
AUCKLAND COMMITTEE ON RACISM & DISCRIMINATION comments to
NZCER about TOSCA and alternative methods of assessment.

This paper is concerned about the way in which tests of scholastic abilities combine with common teaching practices to maintain racism in the education system. We have outlined elsewhere (see Paper 9) our understanding of the term racism, and that understanding is the framework within which we discuss TOSCA and similar tests.

TESTS OF SCHOLASTIC ABILITIES IN GENERAL

There is a substantial literature which demonstrates that, whatever the intention of the constructors, tests of intelligence are racist in the sense of our understanding. First, they rely on the constructor's definition of what is worth knowing, what best represents "intelligence". Second, the items used often include elements that are culture specific, and while some of the cruder examples have typically been eliminated, the majority have not. Thirdly, constructors accept differential outcomes as legitimate. An example is provided in the TOSCA Manual where it was decided that the test would: "predict educational performance in the school system (as it is now) and should reflect the verbal bias of instruction," (e, iii), p19). Thus while items are juggled so there is no overall difference in the average scores of male and female students, the same is not done for ethnic groups or social classes. This decision is justified by pointing to the school system in which the students learn, and saying, in effect, we spread these people in the same way that the school does so we must be measuring the same thing. Expressed as baldly as this the argument clearly fails to recognise either the racism of the school system or the reciprocal links between the testing and the outcomes being used in justification.

There is also a literature concerning the use of such tests, of which Jane Mercer's Labelling the Mentally Retarded (1972) remains one of the most detailed, that shows that the use of these tests discriminates against students who do not belong to the dominant cultural stream. The outcomes are mediated in a variety of ways: teacher expectations, curtailed educational opportunities, and student self-perceptions are important examples. Most importantly, for an evaluation of TOSCA, it is the professionals, usually white, who are consistently in favour of the use of such tests. Often their arguments are presented as representing the children's best interest, as with "slow learner" classes, but the evidence is that the outcomes are not, in fact, favourable to most children. Finally, the opposition to the use of these tests has been strongest among those who have seen their children educationally damaged by their use. That says very strongly that no assessment of these tests and their use can be regarded as adequate if it does not hear the voices of those who have been victimised by their use.

TOSCA: TEST OF SCHOLASTIC ABILITIES

In 1976 the Auckland Committee On Racism & Discrimination expressed strong opposition to the proposed development of TOSCA. That opposition was based on the following grounds:-

1. That such tests are not valid measures of the abilities of students who are not members of the dominant culture.
2. That the tests distort the relationship between teachers and students by creating self-fulfilling prophecies of success and

failure that are substantially sustained by teacher expectations.

3. That the tests are validated by reference to academic success, and act to maintain the reduction of options for those designated 'less able' by the test.
4. Those who were encouraging the development and use of such tests were the 'successful' who had benefited from an education that included such tests.

It was our contention in 1976, when we responded to the NZCER Intelligence Testing Survey, that tests such as TOSCA are a significant part of the racism of the education system. We argued the case with Mr Neil Reid (30/4/76 & 9/8/76) and received from him (7/10/76) and Mr John Watson (13/5/76) reassurances that the NZCER was not only aware of the issues but was sensitive to racial and discriminatory practices. We were also told (Reid, 7/10/76) that we had not been alone in raising the issue.

It is a matter of fact, opposition and sensitivity notwithstanding, that TOSCA was developed. At least five years' work from the Council's Test Development Division produced a group administration test of general scholastic ability that correlates well with success in the present school system and which definitely discriminates against Maori pupils (Nash, 1984). In 1976 we argued on the basis of overseas experience and research with the use of intelligence tests in schools. Now we are drawing on a substantial survey of the use and effects of TOSCA (McCreanor, 1984).

McCreanor's survey of 83 of the 92 schools in the Auckland region showed that 75% practiced some form of streaming of their third form intake. 84% used standardised tests of ability for streamed or unstreamed class placement. Already, in 1983, 40 schools used TOSCA. He collected data from the schools which used TOSCA: his sample included 6525 pupils. Apparently, even as soon as that after its release, the test was being used on a considerable proportion of New Zealand pupils. Overall nearly 20% of the variance in TOSCA scores was predictable on the basis of Ethnicity and Social Class. The detailed analysis showed that both Maori and Other Polynesian pupils were grossly disadvantaged by the test. This can be seen by inspecting the decile distribution. Deciles divide the original group, the ones on whom the test was tried out, into ten equal groups. If the test is working for other groups of students in the way it did for the original group there will be 10% of them in each decile. For the European students there are too few in the lowest decile, but otherwise, despite some unevenness, they are reasonably close to the theoretical mark. In contrast 41% of the Maori and 56% of the Pacific Island students were in the first two deciles (which should account for only 20%). There was also a social class gradient such that pupils from working class families were similarly, if less strongly, disadvantaged. The results were similar to the U.S. Army Alpha test of World War I, with zeros being disproportionately common. It is important to understand that these gross discrepancies remain in the outcomes because of the decision to have the test reflect current education advantage. The mix of items included was intended to produce no reliable difference between the sexes: McCreanor's survey shows that this social objective was achieved. We challenge the appropriateness of the criteria used in item selection and point out that a test which discriminated against Pakeha pupils in this way would not have been acceptable.

McCreanor showed that the effect of administration in schools was to increase

differentials compared to studies that appear to have used more professional administration, smaller numbers and situations of lesser anxiety. If that were to be substantiated in further studies it would be another strong reason for withdrawing TOSCA and similar general abilities tests from schools.

From McCreanor's survey we see that the primary use of the tests is to stream pupils, a practice of dubious educational value. The major support for such testing comes from teachers who have been inadequately trained to teach broad ability classes effectively. It is important to remember that those consulted by NZCER before the development of TOSCA were "primary, intermediate and secondary teachers, guidance counsellors, inspectors and teachers college lecturers." (TOSCA manual, p 19). These are the successful of the education system and they are overwhelmingly Pakeha: we believe that they saw the value of the proposed test from that perspective and that that contributed to the racist outcome.

In considering the uses to which the tests are put it is not sufficient to focus on the inadequacies in teacher training that maintain the support for streaming. McCreanor's survey showed that some schools used standardised tests for placements in unstreamed classes. It is necessary to look at the assumptions. Aptitude tests like TOSCA, and the practice of streaming, assume that each child has a certain amount of teachability and that teachers are most effective when they have the child's measure. This ignores all the data about the role of feedback in learning and the impact of teacher expectations, often fuelled by test results, on classroom practice.

LEARNING AND ACHIEVING

The Second Report of the Committee of Inquiry into Curriculum, Assessment and Qualifications in Forms 5 to 7, Learning and Achieving, makes important points about assessment that relate to tests of scholastic abilities. Having said that assessment should be an integral part of curriculum planning (p 14) and that students should develop skills in self-assessment (p 56) the committee appears to assume that achievement based assessment is a relatively straightforward process. As almost all teachers were trained on a form of norm-referenced assessment, we believe it is essential that organisations like the Council provide assistance and training in this new development. Scholastic ability testing must be geared to provide clear information about the level of skill or comprehension attained. This implies a wider range of tests, a complete rejection of norm-referencing as a mechanism for scaling test items and training of teaching staff to develop tests that can be used relatively widely.

There are also important implications for the use of existing scholastic abilities tests. The practice of streaming, of identifying "bright", "slow" or "normal" pupils to be taught in separate bands, along with national rank-ordering examinations is discouraged. The evidence too clearly shows that such classifications prejudice pupils' educational opportunities. Again, the need will be for tests that are more specific. There is little value in global ability ratings that do not identify what the individual can or cannot do. TOSCA is clearly inadequate to meet this need, having been devised to "present a single total score as a summary of test performance," (TOSCA Manual, p 19 (e) xi)). It does not provide data that can be reliably diagnostic. The theory of mental abilities on which TOSCA is said to be based is not linked either with specific scholastic abilities or with identified sets of items in a way that might permit some diagnosis of strengths or weaknesses in student

performance. Even if this were changed, the results would be lacking the necessary reliability and the preferred alternative would remain sets of smaller, content and/or skill specific, interlocking tests.

IMPLICATIONS FOR NZCER

There is a clear case for the withdrawal of TOSCA and other category A, norm-referenced tests of scholastic ability. Serious consideration must be given to restricting the use of other norm-referenced tests of intelligence/ability.

The Second Report of the Committee of Inquiry into Curriculum, Assessment and Qualifications in Forms 5 to 7 identifies some important tasks in which the Test Development Division of the NZCER should be involved. First, the actual development of criterion-referenced (p 52) and achievement-based (p 57-9) tests for particular subjects and content areas that are appropriate for this country. We commented earlier that the report fails to recognise the work required to create such tests. The resources of the Test Development Division should be directed to these tasks for the next ten years. This requires a reassessment of the staffing of the division, not only in terms of their expertise but also their commitment to and competence with these new ways of 'testing'. We believe that their commitment to a bicultural way of working is particularly important at this time.

Further, before test development in the new mode can proceed, there will need to be retraining or reorienting towards criterion referenced test development. This need not be exhaustive, but must ensure that the staff have sufficient background to carry through what will be a new set of processes, effectively.

Second, and in parallel with the tasks outlined above, will be work with teachers through the department and the teacher organisations. Here the task will be to enable teachers to use the new test effectively and, before too long, to contribute to the creation of further tests that are needed. Teachers will only make such changes if their initial experience of the new forms of assessment is positive. Such positive experience will come from teachers being competent at the mechanical processes of administration and feeling confident about building the outcomes into their teaching practices.

Thirdly, there is a need for Maori and other Polynesian staff who have both technical skills and a firm foundation in their other-than-Pakeha culture. The NZCER should explore ways in which suitable candidates might gain the necessary skills to fill appropriate posts.

Finally, all staff need to have a strong anti-racism commitment and work to a clearly-stated code: TOSCA has showed very clearly that sensitivity is not sufficient.

Prepared by: Margaret Arthur
David Earle
Ray Nairn

For ACORD

FURTHER READING

McCreanor, Timothy, TOSCA: A Survey of its Use and Implications: thesis précis
Nash, Roy, "Why TOSCA should not be used to stream Maori children".