SOME PRACTICAL AND THEORETICAL PROBLEMS OF GENERAL ABILITY TESTING AT THE AFRICAN STANDARD SIX LEVEL IN SOURTHERN RHODESIA.

by

S. H. Irvine.

(Lecturer in Education, University College of Rhodesia and Nyasaland and Acting Director, Rhodes-Livingstone Institute).

The testing of abilities in Africa, until after the Second World War, was largely comparative in nature and theoretical in its approach. It was designed to illustrate differences between ethnic groups and their respective cultures. Applied research, where emphasis was placed on the relationship of the individual to his peers has been the result of the need to classify and use African labour in primary and secondary industry and lately, to cope with the problems of selecting from a large number of primary school leavers, those who could most benefit from further schooling.

It is this particular problem of secondary school selection in Rhodesia that taxes the Mental Ability Survey being conducted by the University College of Rhodesia and Nyasaland. This paper sets down some of the results of the work so far, comments upon their practical significance and perhaps adds something to the long line of theory that has developed out of the early anthropological approaches to testing.

The project began in discussions between the Institute of Education and Department of African Education conducted late in 1961. In March 1962 funds were raised* and approval given for a pilot study, completed in July, 1962, and for a large scale exercise designed to examine the structure of the traditional qualifying examination (henceforth Q), its predictive efficiency, and to assess the contribution of tests of general ability to the selection procedure which had, before this, been based on the raw score totals of the Q. The work, it was agreed, should be purely experimental. At present, the large-scale field work is complete, and the data are being processed for computation. There remains the systematic validation of the measures used. The remarks that follow are based on the Pilot Study and the 1962 Q results for the sample and the experience gained in administering the tests to over 2,000 children. The practical problems of carrying out this exercise will be discussed first.

The Pilot Study was planned to help choose from a battery of tests those that could best be used on a larger scale. Accordingly, a short list of tests was drawn up with reference to the Report issued by C.C.T.A. (1962). They included Raven's Progressive Matrices (1938), Morrisby's Compound Series and Numerical Ability, all non-verbal tests; and N.I.P.R.+ Spiral Nines and Mental Alertness, which are verbal. From these, it was decided on the basis of the results shown in Appendix A to persist with Raven's Progressive Matrices and N.I.P.R. Spiral Nines, and to add to these an untried English achievement series, the Vocabulary, Comprehension and Spelling Tests of the N.I.P.R. Normal Battery.

---

* From the U.C.R.N. Research Fund, Anglo-American and R.S.T. Copper Companies and the Department of African Education, who released Mr. E. Ratcliffe-Smith to assist in the field.

* N.I.P.R. National Institute of Personnel Research, Johannesburg.

Even in this small exercise the problems of Sampling, Administration and of finding suitable criteria for assessing validity were encountered sharply - and they were to be met again in the larger study. Let us take each of these in turn.

The pilot study sample was wholly urban, and for a reason that is only apparent when one understands that the policy of the African Education Department in Southern Rhodesia has been to provide, as a social measure, places for as many of the urban children who present themselves for schooling as is possible. This means that in the urban Standard 6 classes the range of ability is wider than anywhere else in the Territory, because of the relatively unselected nature of the pupils in them. Hence it was postulated that the range of scores on the tests would include those likely to be made by pupils in all types of schools, except perhaps, at the upper end of the range, where only the most selective mission boarding schools might differ. The samplewas then chosen at random from the school rolls of the five schools in Harare township, Salisbury, every third child being selected. In one school however, as the children were unstreamed, a single class was taken, a decision that had important consequences, since it was realised that here was an opportunity to study the effect of variance in marking and teachers' assessments in one class where the processes were handled by one man only, compared with the other schools, where three or even four teachers were concerned in the marking of school examinations and in assessing the groups of children drawn at random from the school population. The criterion finally used for concurrent validity was the school mock Q papers in mid-year, and which previous probes in 1961 and 1962 had shown to have validities of .78 and .60 respectively, with first term marks in Form I in one Government school. The sample figure levelled out at 204 children.

The larger study took place in Mashonaland, the central region of Southern Rhodesia, since it had 12 secondary schools, including urban Junior secondary day schools and thus offers a large follow-up population. The sampling model, then, was based on the school as the unit. It was known that wide differences in types of school existed through the region. Each, it seemed, had to be represented in the final analysis. Geographically, too, the schools were widespread, introducing a second important factor. Hence the schools were classified by District Commissioners' areas, graded by the Inspectorate, and a sample was chosen with reference to area and grade, completing a group of 31 schools and, finally, 1,819 pupils, representing one third of schools presenting candidates for the "Q" examination in 1962.

The acceptance of the widely held principle of the interdependence of general ability and school achievement for the realisation of both, synthesised admirably by Vernon (1960) made this sampling model a necessity, and the decision is further strengthened by the evidence of Davis (1928) and Schmidt (1960) who demonstrate the positive relationship between length of schooling and intelligence that scores with groups of American Negroes and Natal Indians respectively. Furthermore, the researches of Hartson (1936), Sutherland (1951) and Alon-Bakaliar (1952) show that the type of school, its curricular emphasis and the heterogeneity of its pupils, may all exert their influence on performance in achievement and intelligence tests. If such is the case in societies where the extra-mural culture reinforces the learning habits and conceptual manipulation required for high academic aptitude and skill, how much more must it be so in Africa, where the school is virtually the only consistent vehicle for transmission of the exercises required to foster a particular type of thinking, especially in a child population whose parents are largely illiterate.

The clinching piece of evidence for this decision was found when the norms for the N.I.P.R. Mental Alertness Test, that had been tried out in South Africa, were compared with the Rhodesian Norms that the Survey had gathered in the Urban Pilot Survey. The test is largely one of verbal adaptability. By this is meant the capacity to manipulate relational thinking in English or a similar Indo-European language. For African Children this is a complex mental task, since English itself may condition the type of verbal concept required, and the exact relationship of language structure to the nature of thought itself may differ (Biesheuvel, 1958) from Indo-European to Bantu Languages. Nevertheless, verbal adaptability tests combine ability to think relationally in, and skill in adapting to the conceptual mould of, English. Test scores, it is considered, will then be combinations of general and specific ability, largely qualified by achievement patterns in English. The Harare mean on the Mental Alertness Test was 21.1, compared with the South African Form I rural mean of 20, and urban mean of 23*. Here then, is the most unselected Standard 6 group in Mashonaland, Southern Rhodesia, straddling the highly selected First Form South African group. Again, the South African Fourth Form means on the same test were Rural, 30, and Urban, 34-35 points of raw score. Compared with a First Form group of 70 pupils at a Mission School in Southern Rhodesia, who had an average of 31-32 points.

The Bantu educational policy in South Africa insists that the Vernacular languages must predominate. In Rhodesia, English is taught as soon as possible. These two sets of results, then, may indicate that the English medium teaching in Rhodesia has resulted (a) in a greater ability in English in a shorter time and (b) in a much more flexible manipulation of the type of concept that is necessary for constructive relational thought of the sort demanded by scholastic training in a Western technological environment. The argument by Silvey (1962) that variety of environment was liable to be a more important factor than formal schooling in the development of conceptual flexibility may indeed be true where the standard of teaching, the curriculum itself, and the time devoted to various aspects of the curriculum are more or less homogenous. Yet one might suppose that the variety of environmental surroundings would be greater in South Africa than in Rhodesia, because of its greater industrialisation and more numerous white population; but it is difficult not to conclude that the type of schooling given seems to be as crucial a factor as environment, judged from these comparisons and the evidence of the research cited. The Rhodesian sample, then, for all these reasons, was built round the school, in the belief that our tests may be far more hindsightful (Cf. Schwarz 1961, p. 6) scholastically and environmentally, than they have proved to be in Western societies. How far ahead they will predict, or how foresightful they will prove, is a matter yet to be discovered. Our sampling frame, by virtue of the grading of schools by Inspectors, and by an additional index of teaching efficiency that we have devised on the basis of past Q results, will perhaps permit analysis of covariance in order to assess the effect of a school's quality on the score of the individual in this particular part of Rhodesia.

Administering the tests to the two samples described above was a challenging and exacting task. It was soon realised that the skills required were those of the teacher. Schwarz (1961) has formulated nine principles for the administration and construction of tests for African populations. Some of these are certainly applicable, particularly his first maxim "Test procedures should not presuppose any response as being automatic. Every response should be taught". However, a standard procedure was evolved that was essentially architectonic in nature. As we dealt with each test in turn similarities in

---

* The South African figures are approximate, being taken from the norm sheet kindly given by Mr. A.O.H. Roberts of N.I.P.R.

in recording procedures were emphasised, reinforced and demonstrated, so that by the end of the morning needed to apply the battery, the children did, in fact, respond automatically. Both testers were trained teachers and constant discussion, observation and conference fixed the procedures early in the programme. Instructions were given in English : Individual pupil participation and questions were features of demonstrations. In fact, normal classroom procedures were used, and here we part company with Schwarz who recommends that test procedures should differ sharply from school routines. True, new notions should be presented in an alive, interested and at times dramatic manner, as he suggests, but as teachers, we could see nothing abnormal in this. Maximum cooperation, we found, was not so much a question of presentation, but a question of allaying fear, suspicion and lack of confidence, when it arose, by personal kindness, attention, and face-saving for the pupil in difficulties, by enlisting the aid of the class teacher and getting his assistance at every opportunity and by ensuring that the headmaster personally introduced us to our classes as people to be trusted. An interesting effect, we found, was exerted by Raven's Progressive Matrices. When the children and the teachers saw this test, they accepted it not as a 'foreign' test, but as one in which day-to-day school work could exert no influence, so that they really thought there was no bias in the test situation towards the traditional Q examination. Admiration for this type of test was expressed time and again, and also for the general make up of the other tests, which suggested careful organisation and planning. All these things, small in themselves, but together cumulative, smoothed our way through a difficult period in Southern Rhodesia, following the banning of the Zimbabwe African People's Union in September, by which time we were three weeks into a nine-week field trip that took us 3,000 miles round Mashonaland, and into its furthermost geographical regions, reserves and purchase areas. We used mission centres as bases and radiated to out-schools from these. The urban areas were completed first and this gave us an interesting comparison between the tempo of work in these schools compared with the rural schools. A few comments, merely observational, but related to the general problem of test administration are offered on this. It was found that the rural children tended to work, as a class, at a more or less uniform rate. True, there was little need to hurry, since the tests were generously timed. The tempo in urban areas was far more varied, and this difference we took to reflect radical differences in cultural adaptation. This perhaps, was also a function of the less selected population in the urban schools, a fact that would imply different rates of work, but we deliberately gave a time warning on several occasions and were rewarded by a raising of tempo, or attempts to do this, by even the slowest and apparently least able pupils in urban schools, while in rural schools, there seemed to be a sense of fatalism about it all and little visible change could be observed. Smith's (1948) work with Hebridean children shows how leisurely environment can affect intelligence test responses, so that the maxim that tests for African children should be tests of power rather than speed and its corrollary that tests should not be constructed in such a way as to make differences of approach a handicap, seem to be upheld from our observations.

In case it should be thought that we are advocating general rules for test administration, it must be said that the pilot study investigation taught us caution in this. The Raven Test was tried out in three different ways, in order to measure the effect of the different presentations on groups drawn from the same population. In one, a simple introduction was given to the test and a story concocted about a hole in a shirt that needed patching. In the second and third, a flannelgraph build up of six different examples of test items was demonstrated for each, the only difference being that one group was allowed to go through the test in a single unit of time, and the other have the test broken up into its five sections, timed separately. We eventually decided that the

first method was preferable, since, as the Congo Studies (1958) on the Raven showed, the results of the second and third methods narrowed the standard deviations and decreased the internal consistency of the test (see Appendix B). In addition, the simpler a test is to administer, the more easily it can be used by practising teachers as a local tool, with reference to the norms collected for it by investigation. Silvey (1962) has found, however, that in Uganda it is preferable to use a modification of one of the alternatives, and once again the vastness of Africa illustrates how experiment under local conditions makes something of general principles, however well intentioned they may be, and however well founded they are locally, an absolute necessity.

Our all-over debt to schools was large, particularly in the Pilot Study, where concurrent validities with (a) school examinations, (b) streaming or continuous educability estimates and (c) personal assessment made by the teachers, were all tried. The problem of finding a criterion as Travers (1949) and Dale (1960) point out, was complex. The charge that the criterion had been arranged by the experiments, or that it omitted something vital, which Dale makes against the usual type of validity study, had to be considered most carefully, yet Rodger's (1949) statement, that selection mechanisms had to be "technically sound, administratively convenient and politically defensible" was never far from our preliminary thoughts. At first, all three were tried independently, as Appendix A shows. Then, after agreeing with staffs that the school examinations should remain as the main criterion (mainly on the basis of Rodger's last two criteria) to add the streaming, or continuous educability, estimates to the test battery, since we know that these had been provided before the mock Q examinations had been given, whereas the teachers' personal assessments given after the mock Q, by four out of the five schools sampled, have remarkable and probably spurious relationship to the mock Q. results, and were therefore discarded, since their inclusion would have greatly inflated multiple correlations with the school examinations on which they were based.

The interesting result of multiple correlation between school examinations and the tests used was the variation in these correlations from school to school. They varied from .589 to .870, with pooled correlation* of .611. The highest correlation occurrred in the school where only one person had dealt with marking and assessing. In other words, where teachers marked their own class examinations, as they were reported to have done in all the schools where the pupils were randomly selected from three or four streams, the correlation of the tests with the criterion could have suffered because of possible error variance in the criterion itself. Another possibility is that the school examinations tested different things in different schools. The closer the examinations came to testing what our tests tested, the closer the correlation between them and our tests' scores would come to unity, if the tests were functioning correctly and accurately. Of course both variance in marking and criterion instability could act together to cause differences in validity from school to school. Silvey's experience (1962, p.8) of correlation between school examinations and Raven may be a function of this. While this is so, it would indicate differences in approach from school to school that we have already noted and hoped to take account of in sampling to see whether the final system of selection will have to become statistically sensitive to such differences.

---

* Not altogether accurate, because independent school examinations were not commonly scaled.

The reason, however, may lie in another direction altogether, as was pointed out by Mr. A.O.H. Roberts of N.I.P.R. in a communication to the writer. The multiple correlations with the criterion depend on the validities of the individual measures used. Adopting postulated population values for the validities of individual tests, it was found that only in school 5 (see Appendix B. Table 2.) did two test validities fall outside the 95 per cent confidence limits for the population values. In other words, it is perfectly possible that the differences in multiple correlation were due to chance fluctuations within the sample of which the five sub-groups formed a part. Enough has been done, however, to show that concurrent validity in the pilot study varied from school to school, and varied enough within the limits of chance, to make pooled predictions difficult. Our conclusions must, it seems, be in line with those of Wantman (1962 a.) who states that tests are more valid from one school to another than from a number of schools to another, unless one has time to adopt, in practical exercise, the statistical sophistication of Tucker (1960) who makes use of factor analysis to provide a model for within schools correlation of a criterion with variables from a number of schools whose grades and marks are not comparable.

Before leaving the question of validity, it must be noted that the addition of the streaming or continuous educability estimate greatly increased the multiple correlation with school examinations. It can be argued that examination results are a function of streaming, and this may be true. However, no scaling of these estimates was attempted and they carry with them the errors that are made by schools, so that considerable overlap may occur between performance in one class and another. Moreover, there was enough of an improvement to allow the conclusion that teachers do know the capabilities of their children, and in accordance with the classical study of McLelland (1942) and the later confirmation of Emmett and Wilmut (1952), Bosomworth (1953), Vernon (1957), Yates and Pidgeon (1957) and especially Wantman's (1961) report of his work in Malaya, it was decided to get such an estimate from each class teacher in the field study and feed it in as a possibly important factor for prediction of success. The last of the practical problems concerns the traditional Q examination. The terms of the Survey included an investigation into the fitness of the Q to predict success in Secondary schools. At present, very little criterion evidence is to hand except the preliminary probes already reported. Some internal evidence has, however, come to hand from the collation of 1962 Q results from the sample.

TABLE 1*

Summary of Q Examination Sub-Test Scores
at various Percentile Points

S. Rhodesia 1962. (N = 1819)

Subject	Pass Possible and %		Median 50th% ile		65th % ile		75th % ile		85th % ile		95th % ile
	a	b	c		d		e		f		g
				3		2		2		5	
English 1.	55 (28)	34%	25	→	28	→	30	→	32	→	37
				4		2		2		4	
English 2.	70 (35)	92%	45	→	49	→	51	→	53	→	57
				7		5		5		8	
Arithmetic 1	100 (50)	85%	68	→	75	→	80	→	85	→	93
				8		6		7		10	
Arithmetic 2	100 (50)	31%	41	→	49	→	55	→	62	→	72
				2.5		1		1		2	
Mental	25 (13)	72%	15.5	→	18	→	19	→	20	→	22
				3		2		2		3	
H.C.A.	50 (25)	80%	31	→	34	→	36	→	38	→	41
				3		1		2		2	
G.N.S.	50 (25)	86%	32	→	35	→	36	→	38	→	40
Totals				50.5		19		21		34	
				23		10		15		21	
G. Total Q.	400 200	85%	240	→	253	→	263	→	278	→	299

* The results are in columns as follows:

(a) Possible mark
(b) Arbitrary pass mark and percentage of sample bettering this.
(c) Median
(d) 65th percentile
(e) 75th percentile
(f) 85th percentile
(g) 95th percentile

Figures above arrows indicate differences in raw scores between the percentage points.

H.C.A. = History - Current Affairs.
G.N.S. = Geography - Nature Study.

Totals - sum of percentile Differences in raw scores for each test.

At present, candidates may fail the Q in one of three ways: (a) a failure to obtain half marks in the total of the two English examinations. (b) a failure to obtain half marks in total gained from the average of the two Arithmetic tests, plus the Mental Arithmetic, and (c) a failure to obtain, even if (a) and (b) are fulfilled, a total of 200 marks out of 400. From Column b in Table 1, it will be seen that the two most difficult papers were English (34% above pass) 1 and Arithmetic 2 (31% above pass.) The others show that the pass mark, arbitrarily set, could be [illegible] attained by all but the poorest of pupils. Even so, the averaging of the two English and two Arithmetic papers would remove most of the sting from the difficult tests, since the difficult tests happen to be paired with two of the easiest papers. As a pass-or-fail test, set at the end of eight years of education, the Q may serve its purpose. The Q, may, on these figures, show that the teachers in the schools sampled have reached a high degree of proficiency in drilling their pupils to pass. English 2, (Comprehension, Grammar, etc.) Arithmetic 1 (Mechanical) and the two 'General' Papers (H.C.A. & G.N.S.) have been taught alarmingly well.

However, this test has been used since 1960 as a selection device for the secondary schools, again with arbitrary cut-off points, using the Global Q marks as the yardstick. The point to note in Table 1 is the way in which the examination total spreads the group. Only after the 85th percentile does the examination as a whole start to show signs that it has proved that it can spread the pupils over a large range of marks. Between the 65th and 75th percentile, only 10 points of raw score, if these marks were to represent the whole of Southern Rhodesia, would separate 10 per cent of the total of candidates presented (some 16,000) and this after examinations lasting a total of over 10 hours. As some 25 per cent of the total find post-primary places either in boarding or day secondaries or in vocational training and the border zone is critical, it is precisely in the 65th - 75th percentile area that the examination should demonstrate most clearly. This, on the Global Q mark, is where it is weakest. It should be noted, however, that if the Arithmetic and English tests were not averaged before being totalled as they are at present, the indication given by the addition of single test score differences between the percentiles, is that the range might increase. This is not necessarily so, however, since the total range and variance will only increase if the individual tests tend to put the same pupils in the same order - and our results do not, at this stage, indicate if they do this.

Our problem here, however, is not just one of increasing discriminative power at crucial ranges, but of educating a public into realising that the day of arbitrary cut-off points must come to an end soon. In Mashonaland, in 1962, only 5 per cent of our sample (which is representative) gained a Grade 1 pass (over 300 marks) compared with 15.6 and 12.3 per cent in 1960 and 1961. A Grade 1 pass has, before this, been the pre-requisite for entry into a secondary school offering four years of education to entrant. This year, headmasters, staffs and children have all been surprised and upset. The situation is entirely due to the system of arbitrary cut-off points that take no account of annual variations in population ability and in examination difficulty. This type of result can only be expected as a more and more unselected group attains Standard 6 as the years pass. Accusations will be many and the Grade system must become flexible, rather than rigid, and depend on the systematic gathering of normative data before decisions for grades of passes are made. At present, the data collected from the sample are being transformed into Standard Scores and Stanines, so that each test will have equal weight, and that each test will be comparable with any other. This is as good a base-line as any from which to start more complex procedures.

This section on the practical problems of the Mental Ability Survey in Southern Rhodesia has dealt with sampling, test administration, the choosing of criteria and the structure of the Q examination. A shift in emphasis to the theoretical hypotheses that have come out of the work so far will occupy the last part of this paper.

The question of what, in fact, we have been testing in this type of study has already been explored by various writers, particularly Biesheuvel. Although many reasons have been advanced for not comparing test scores made by Europeans and Africans on the same tests, not the least of which has been the impossibility of equating environment for the groups under comparison, no writer has yet taken what seems to be the logical step of stating that the meaning of test scores for African groups, at present, is distinctly different from that which can be given to similar test scores for groups in western societies. This move then, makes possible a meaning of intelligence that is neither intelligence A, genetic potential, nor intelligence B, "the actual

level of comprehension, learning and problem solving in this (western) culture" (Hebb, 1949, p.299), which is a condition of environment, nor (Vernon, 1957) intelligence C, a sample of B given by any test. The meaning of African intelligence may, at this stage of the development of its peoples, be different from any of these, because of several sharply conditioning factors requiring an interpretation of test scores that will not fit into the above categories. This meaning we may designate Intelligence X.

Already in this paper, an example has been given of the difference in test scores between South African and Rhodesian Sunday School children. This was attributed largely to a variation in schooling. The individual studies of Gordon (1924) and Asher (1935) illustrated early in the history of intelligence testing that educationally deprived sub-groups within western cultures did less well than those receiving a normal education. The South African/Rhodesian results, of course, raise the question of what is a normal education. There is no answer to this, since children in any country receive the education that is laid down for them by the law. African education in South Africa, however, in terms of the results of this test of verbal adaptability, shows up badly with education in Rhodesia. The South African groups could be said to be less intelligent, for all their superior education. The criticism to this argument may be that the test of Mental Alertness is more of an attainment test than an intelligence test. Vernon (1957) argues that the distinction between the two types of test is often blurred and, in fact, the tests are the better for being so. But the critics have a point, and a valuable one, for our argument. Scores on tests of Mental Alertness may indeed, in African societies, be far more tests of attainment than intelligence. In other words, the percentage of variance due to genetic potential may be considerably less than it is in Western societies. The 80% genetic and 20% environmental ratio so often quoted, may require drastic revision in Africa, simply because of a cultural anomaly. Anthropologists and psychologists have been quick to point out differences between groups in their social organisation. However, there has been general agreement about the homogeneity of the lives of rural African people growing up in comparable surroundings (Schwarz, 1961), the communal restrictions on individually centred intellectual activities, (Winterbottom, 1948, p. 56) and the general agreement that African societies, considered together, far more resemble each other than they do European societies (Biesheuvel, 1952, b). The great difference, however, within African society, is the degree of acculturation of the individual - the degree to which he has been exposed to Western habits, values, and learning on the one hand and on the other, the extent to which he can command a degree of schematical flexibility with which to absorb, integrate and keep conceptually mobile all the skills he requires for successful participation in conversation, practical and academic skills, though at a pre-determined Western level. These two variables, cultural exposure and schematical flexibility, of which the latter may be a function of the former, are the crux of mental testing in Africa. Our tests may be tests of attainment, not just in the sense of acquiring school drills and skills, but of attainment in a wider social sense, whose exact contribution to the test score is unkown, because of the heterogeneity of the population in their opportunity to acquire such background, when education is not, as it is in Western societies, universal between the ages of six and sixteen. Until we know the precise contribution of environment to a test score which is considered to be derived from a test of intelligence according to its Western precepts, we cannot

argue that we are estimating Intelligence B, as Hebb would seem to think, with any degree of accuracy that will predict, as tests of Intelligence B do in Western societies, with anything like the effectiveness of intelligence tests used on Europeans.

For European children educational background, language, opportunity for conceptual growth, and, even, environment are far more tram-lined to reading and school-readiness, and individual genetic potential can exercise a much more weighty effect on test scores than in Africa, where censor mechanisms of an economic, social and linguistic nature may act indiscriminately on genetic potential to such an extent as to blanket, for the individual, its effect on test scores.

There are several ways in which, at present, environmental factors could act permanently on individual African genetic potential. Schooling, as was said, is one of them. It is perhaps the most interesting point of all investigations comparing non-European with European scores on the same tests that the gaps are narrowest at the beginning of schooling and widest at the end of it. Shuey's (1958) catalogue of results illustrates this most clearly. Tests do not measure the same things at different stages in development, and the gaps in score do not, therefore, mean the same thing nor are they necessarily equatable in size. It could be that tests in the early stages of development do not demand skills of a high verbal order, whereas at the end of schooling they do, since it is precisely this type of skill that has most to offer in a highly complex industrial civilisation.

The crux may indeed lie, in later life, in the type of language which has been handed on in a particular culture. Biesheuvel (1962 a) sums up this argument as follows :-

> "The characteristics of a language spoken by a people therefore determine to a very large extent the nature of their thought processes, particularly the extent to which these can function at the conceptual level ....the manner in which the problems (in tests) are formulated and the thought processes required for their solution ... are still characteristic of the Indo-European languages".

Although English is taught as a second language almost from the start of schooling in Central Africa, it does not become fully the medium until at least two, more often three or four years of schooling. And here another factor enters into the discussion. The work of Hebb and Piaget both give weight to the theory that certain stages of readiness in the behaviour of an organism demand certain types of stimuli for the full maturity of the organism's potential. Should these stimuli be withheld, then atropy and rigidity in behaviour could well-nigh result. If, then, Biesheuvel's (1958) comment that "performance also depends on a process of logical reasoning which requires verbal concepts that are not characteristic of African languages" is so, then the combination of late familiarity with the "conceptualistics" of Indo-European languages and the lack of certain elements of these concepts in Central Bantu languages would produce a double barrier to the African child's mental growth in a Western society that might never be overcome. A result from the Rhodesian Pilot Study encourages this hypothesis. At present there is a small, persistent, but statistically insignificant negative correlation between age and test scores. We need further evidence to show whether or not this correlation is a chance one or not; but if it is not, then late entry to school for the older children could mean

that they have, in fact, passed the optimum period for mental stimulus on entry and that the "general level of complexity and flexibility" of their schemata (Vernon, 1960, p. 39) had been permanently lowered. The much more complex problem of investigating the effect of early concept development in the African vernacular, with its relation to the type of concept formed by the European child, has yet to be tackled.

African society at present, then, may hand on a language that has gaps in it for the purpose of early forming of schemata that our tests take for granted; and the policy of its schooling may be such as to make the start to filling these through familiarity with a European language too long delayed for many. Is it possible, therefore, to place the same meaning on the word intelligence when we have tested African children as when we test European children ? Can we isolate the variance due to these two factors to such an extent as to make the rank order of the individuals tested as precise as it is in European schools ? And if we did, would the result in terms of rank order mean the same thing for both groups? At present the answer to these questions is, surely, no. Hence, although Nissen et. al. (1935) and Biesheuvel (1943, 1952, a,b,c,) and others have used the term 'adaptability' rather than intelligence, in order to ensure that comparisons between European and African test scores were not erroneously made, the general nature of the term seems to hide the point that all types of mental activity feed back to genetic potential, Intelligence A, and in order to preserve that link, specialised meanings of Intelligence B, such as Intelligence X, are preferable. Biesheuvel, in discussion with the writer, has come back repeatedly to the universality of g, the general factor. His own work (1952 (c), 1954) with the 'General adaptability' Battery showed that a general factor, accounts for 45% of the variance in scores, when the Battery was factorised. Vernon (1961) reports* a two factor result from tests given to African recruits, one of which was general adaptability to the test situation, the other quite close to 'g'. While we have no wish to quarrel with the extent to which 'g' might be universal in any test devised in Western societies to predict academic or near-academic success, wherever it happened to be administered, the variance due to test situation adaptability in Western cultures is minimised, for all the reasons given above, just as it is maximised for African children. The score we get from these tests, then, must be a composite which at present we have broken down only in hypothesis - and it could further be broken down in terms of environmental rigidity and simplicity (Cf. Biesheuvel and Liddicoat, 1959, Vernon 1960, p. 39), and so on.

'Adaptability', then, is the general term for a whole series of complex variables involving several distinct types of 'adaptability' that together produce 'X' scores on the type of test we are using at present. Intelligence X, then, is not Intelligence A, nor Intelligence B, nor, indeed, Intelligence C. It is a global composite which requires a special meaning of its own, until the precise relationship of these scores to what we consider is intelligence B has been worked out.

---

* Noted by Silvey (1962)

APPENDIX A.

CORRELATIONS OF TEST BATTERY WITH SCHOOL EXAMINATIONS STREAMING AND ASSESSMENTS.

(N = 204)

SCHOOL	CRITERIA	RAVEN	M. Al.	C. S.	N. A.	S. N.
1.	S. Exams.	.302	.523	.229	.445	.337
	Streaming	.343	.400	.297	.539	.258
	Assessments	-	-	-	-	-
2.	S. Exams.	.359	.477	.239	.231	.302
	Streaming	.367	.566	.347	.247	.443
	Assessments	.239	.393	.194	.177	.292
3.	S. Exams.	.537	.631	.169	.463	.641
	Streaming	.613	.561	.177	.392	.565
	Assessments	.270	.352	.010	.161	.481
4.	S. Exams.	.468	.594	.498	.463	.381
	Streaming	.397	.140	.478	.291	.234
	Assessments	.212	.419	.418	.290	.204
5.	S. Exams.	.657	.602	.332	.740	.793
	Streaming	.800	.792	.568	.906	.751
	Assessments	.503	.578	.288	.596	.701
All Schools	S. Exams.	.456	.559	.276	.448	.483
	Streaming	.497	.499	.354	.465	.445

Abbreviations used in the tables are as follows:
M = mean; S.D. = Standard Deviation; R_{tt} = Kuder-Richardson estimate of consistency; S. Exams = School Examinations: Stream = Streaming; Raven = Raven's Progressive Matrices; M.AL. = N.I.P.R. Mental Alertness; C.S. = Compound Series; N.A. = Numerical Ability and S.N. = N.I.P.R. Spiral Nines.

APPENDIX B.

TABLE 1.

MEANS STANDARD DEVIATIONS AND. ESTIMATES OF RELIABILITY.

TEST	N	M	S.D.	R_{tt}
Raven (1)	89	24.78	11.23	.896
Raven (2)	49	29.96	7.17	.721
Raven (3)	79	26.94	7.83	.767
Mental Alertness	204	21.11	6.88	.741
Compound Series	204	20.00	6.87	.729
Number Ability	204	44.75	15.93	-
Spiral Nines	204	36.53	7.09	.686

Table 1 shows the mean and standard deviations of all tests used in the battery, and estimates, except for Numerical Ability (N.A.) of internal consistency.

It will be noted that all tests show fair reliability, although the reliability for Mental Alertness and Spiral Nines is worth considerably more than that for Morrisby's Compound Series, which is a homogeneous test, while the former are not. Raven, under testing condition 1, which had no coaching and a time limit of 30 minutes, gave the biggest standard deviation and the greatest reliability. The other two conditions served only to reduce the spread of scores and the test's consistency.

One can conclude, therefore, that the Morrisby results are scarcely satisfactory in the Rhodesian situation, but that the Raven results under condition 1 are comparable with those of the Congo and encourage its further use. Either of the two N.I.P.R. tests could be used with reasonable satisfaction, while the Numerical Ability Test shows qualities of spreading the group as a whole.

TABLE 2.

Validities, Beta-weights, Multiple R and postulated Population Coefficients for Pilot Study Battery.

School	N		1 Ravens	2 Mental Alertn.	3 Comp. Series	4 Number Ability	5 Spiral Nines	M Multiple R
1	46	Validity	.302	.523	.229	.445	.337	.584
		Weight	.040	.383	.115	.260	-.039	
2	38	Validity	.359	.477	.239	.231	.302	.522
		Weight	.201	.534	-.182	-.229	.140	
3	46	Validity	.537	.631	.169	.463	.641	.699
		Weight	.105	.315	-.132	.063	.353	
4	38	Validity	.468	.594	.498	.463	.381	.730
		Weight	.281	.349	.359	.074	-.048	
5	36	Validity	.657	.602	.332	.740*	.793*	.870
		Weight	.200	.057	-.129	.142	.501	
All	204	Validity	.456	.559	.276	.448	.483	.611
		Weight	.156	.322	-.008	.063	.198	
Postulated Population			.45	.55	.30	.45	.50	

* Significantly different from postulated population correlation value at 5 per cent, two-tailed level. All other differences not significant.

REFERENCES.

1. ALON-BAKALIAR, S.L. (1962) Progressive matrices in comparative testing. Habimukh Vol.xxiv,pp.156-159.

2. ASHER, E.J. (1935) The Inadequacy of Current Intelligence Test for Testing Kentucky Mountain Children. J. Genet. Psychol. Vol.xxxx, pp. 480-486.

3. BIESHEUVEL (1943) African Intelligence, Capetown, South African Institute of Race Relations.

4. BIESHEUVEL, S. (1952a) The Study of African Ability. African Studies Vol. xi, pp.45-58, 105-117.

5. BIESHEUVEL, (1952b) The Study of African Ability. African Studies Vol.xi,pp.105-117.

6. BIESHEUVEL, S. (1952c) Personnel Selection Tests for Africans. S.A. Journal Science Vol.xxxxix pp. 3-12

7. BIESHEUVEL, S. (1954) The Measurement of Occupational Aptitudes in a Multiracial Society. Occup. Psychol. Vol. xxviii, p.4.

8. BIESHEUVEL, S. (1958) The Abilities of Africans, in Africa in Transition. Ed. SMITH, P. London. Max Reinhardt, pp. 30-38.

9. BIESHEUVEL, S. and LIDDICOAT, R. (1959) The Effects of Cultural Factors on Intelligence - Test Performance. J. Nat. Inst. Personnel Res. Vol. viii, pp. 3-14.

10. BOSOMWORTH, G. (1953) The use of teachers' assessments in allocating children to secondary schools. Res. Rev. Durban Vol. iv, pp. 50-53.

11. C.C.T.A. (1962) Meeting on the Construction of Vocational Selection Tests to be used at the End of the Primary or Middle Stages. Lagos, Mimeographed.

12. CENTRE D'ORIENTATION PROFESSIONELLE (1958) Etude Sur Le P.M. 38. (Progressive Matrices De Raven). Leopoldville.

13. CENTRE D'ORIENTATION (1958) Contribution a' L'Etude due Progressive Matrices 1938 de Raven. Leopoldville.

14. DALE, R.R. (1960) Reflections on Research on Allocation to Secondary Education. Brit. J. Educ. Psychol. Vol. xxx, pp. 220-24.

15. DAVIS, R.A. Jr. (1928) Some Relations Between Amount of School Training and Intelligence among Negroes. J. Educ. Psychol. Vol. xix, pp. 127-130.

16. EMMETT, W.G. and WILMUT, F.S. (1952) Re Prediction of School Certificate Performance in Specific Subjects. Brit. J. Educ. Psychol. Vol. xxii, p. 52

17. GORDON, H. (1924) Mental and Scholastic Tests among Retarded Children. London Board Educ. Pamphl. Vol. xxxx, p. 92.

18. HARTSON, L.D. (1936) Does College Training Influence Intelligence? J. Educ. Psychol. Vol. xxvii, pp. 481-491.

19. HEBB, D.O. (1949) The Organisation of Behaviour. New York, John Wiley.

20. McLELLAND, W. (1942) Selection for Secondary Education. London, University of London Press.

21. NISSEN, H.W. et. al (1935) A study of Performance Tests given to a Group of Native African Negro Children. Brit. J. Psychol. Vol. xxv, pp. 308-355.

22. RODGER, A. (1949) Symposium on the Selection of Pupils for Different Types of Secondary Schools. VIII. An Industrial Psychologists' Point of View. Brit. J.Educ. Psychol. Vol. xix, pp. 154-59.

23. SHUEY, R. (1958) The Testing of Negro Intelligence. London Holburn.

24. SCHMIDT, W.H.O. (1960) School and Intelligence. Int. Rev. Educ. Vol. vi, pp. 416-32.

25. SCHWARZ, P. (1961) Aptitude Tests for use in the Developing Nations. Pittsburgh. American Institute of Research.

26. SILVEY, J. (1962) Preliminary Thoughts on Aptitude Testing and Educational Selection in East Africa. E.A.I.S.R. Conference. Mimeographed.

27. SMITH, C.M. (1948) Mental Testing of Hebridean Children in Gaelic and English. London.Univ. of London Press.

28. SOTHERLAND, J. (1951) A Comparison of Pupils Arithmetic Ability in the Secondary School with Their Ability at the Time of Transfer from Primary Schools. Brit. J. Educ. Psychol. Vol.xxi, pp. 3-8.

29. TRAVERS, ROBERT, M.W. (1949) The prediction of achievement. Sch. and Soc. Vol. LXX, pp.293-294.

30. TUCKER, L.R. (1960) Formal Models for a Central Prediction System. (Research Bulletin RB 60-14) Princeton, N.J., E.T.S., pp. 1-79.

31. VERNON, P.E. (Ed.) (1957) Secondary School Selection. London (Methuen).

32. VERNON, P.E. (1960) Intelligence and Intelligence Tests. London, University of London Press.

33. VERNON, P.E. (1961) The Structure of Human Abilities. London, Methuen.

34. WANTMAN, M.J. (1962) Memorandum on Joint Project 18. Lagos C.C.T.A. Mimeographed.

35. WANTMAN, M.J. (1962b) Report on the Malayan Project. Educational and Occupational Selection in West Africa, Ed. Taylor, A. O.U.P., pp. 106-115.

WINTERBOTTOM, J.M. (1948) Can we Measure the African's Intelligence ? Rhodes-Livingstone Journal, Vol.vi, pp. 53-59.

37. YATES, A. and PIDGEON, D.A. (1957) Admission to Grammar Schools. London Newnes.
