19 ᵢ

PERS 338

DEVELOPMENT AND VALIDATION
OF THE HIGH LEVEL FIGURE
CLASSIFICATION TEST

NATIONAL INSTITUTE FOR PERSONNEL RESEARCH
Council for Scientific and Industrial Research

NASIONALE INSTITUUT VIR PERSONEELNAVORSING
Wetenskaplike en Nywerheidnavorsingsraad

ND VALIDATION
EVEL FIGURE
N TEST

NATIONAL INSTITUTE FOR PERSONNEL RESEARCH
COUNCIL FOR SCIENTIFIC AND INDUSTRIAL RESEARCH
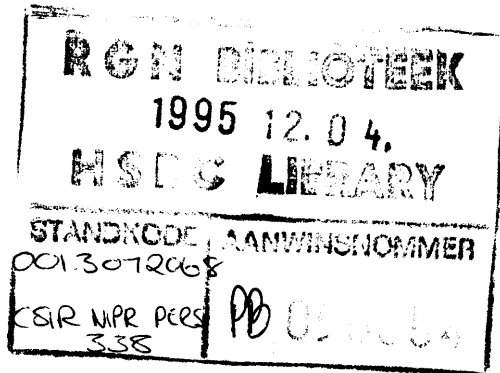
CSIR, NIPR Report PERS 338

A C K N O W L E D G E M E N T S

S U M M A R Y

This report describes the history, construction and validation of the High Level Figure Classification Test. The test was classified by the Test Commission as an 'A' level test in September, 1982.

The High Level Figure Classification Test went through three major stages in its development. In this process the number of items was reduced from 54 to 36 and finally to 24.

Detailed analysis of the High Level Figure Classification Test was carried out on ten samples representative of the population groups for which it is intended. This process is described with particular reference to the main analyses - item analysis and item response evaluation.

The merits of the High Level Figure Classification Test as a selection instrument are discussed.

O P S O M M I N G

Hierdie verslag beskryf die geskiedenis, konstruksie en geldigheidsbepaling van die Hoëvlak Figuurindelingstoets. Die toets is in September 1982 deur die Toetskommissie as 'n 'A'-vlak toets geklassifiseer.

Die Hoëvlak Figuurindelingstoets het deur drie belangrike ontwikkelingstadia gegaan. Tydens hierdie proses is die aantal items verminder van 54 tot 36 en uiteindelik na 24.

Deeglike ontledings van die Hoëvlak Figuurindelingstoets is uitgevoer op tien steekproewe wat verteenwoordigend is van die populasiegroepe waarvoor dit bedoel is. Hierdie proses is beskryf met besondere verwysing na die hoofontledings - itemontleding en itemrespons-evaluasie.

Die voortreflike hoedanighede van die Hoëvlak Figuurindelingstoets as 'n keuringsinstrument is bespreek.

TABLE OF CONTENTS

TABLE OF CONTENTS	PAGE

L I S T   O F   F I G U R E S

L I S T   O F   T A B L E S

L I S T   O F   T A B L E S

1.0 INTRODUCTION

Ever increasing numbers of Blacks are receiving secondary school education.
The surge in the educational level of Blacks over the past 10 years
has created a demand for psychological tests to assess the cognitive
abilities of Blacks with secondary school education.

The NIPR reacted to this demand in 1974 when development began on the
Figure Classification Test, a non-verbal test of concept formation.
The test was published in 1976 (see Taylor, 1976) and was designed for
use in the Standard 5 - 8 range.

The Figure Classification Test was 'pitched' at this educational range
because of the rapidly rising numbers of Blacks who were starting to go
to high school. At that stage, few Blacks reached or went beyond
Standard 8, however (van Rensburg, 1975).

In the past few years, this picture has begun to change. Since 1979,
there have been annually more than 100 000 Blacks in Standard 8 alone
and in 1981, there were 136 000. This should be compared with the
1971 figure of 30 000. Over the 10-year period 1971 - 1981, the number
of Blacks in matric has increased tenfold, from 4 100 to 43 600.
Figures 1 and 2 present graphs of numbers of Blacks in Standard 8
and matric over the 1971 - 1981 period. These statistics were drawn
from the Report of the National Manpower Commission for the period
1 January - 31 December 1981.

THOUSANDS   NUMBERS OF BLACKS IN STANDARD 8 OVER THE PERIOD
1971-1981
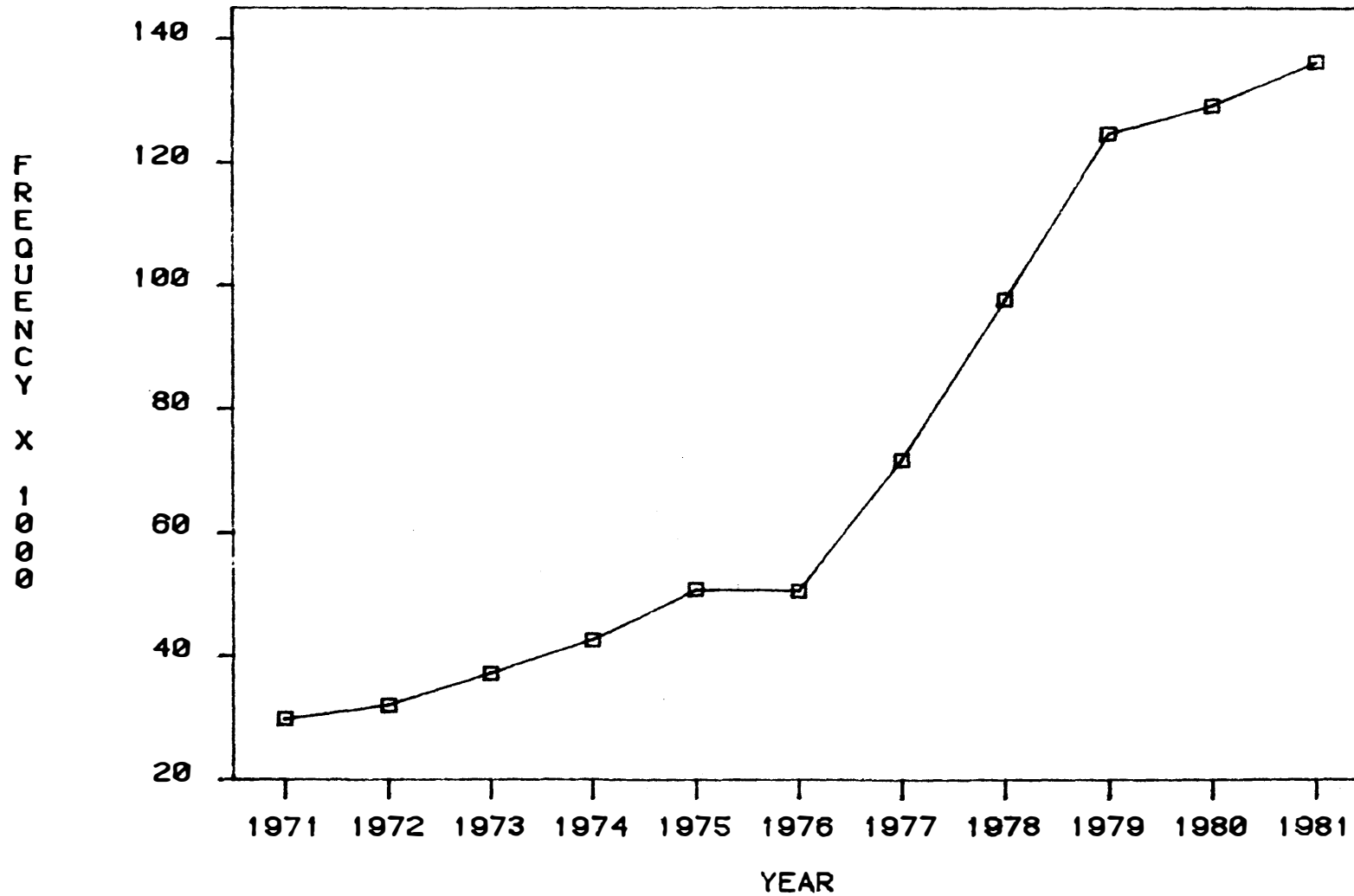
Figure 1

THOUSANDS   NUMBERS OF BLACKS IN STANDARD 10 OVER THE PERIOD
1971-1981
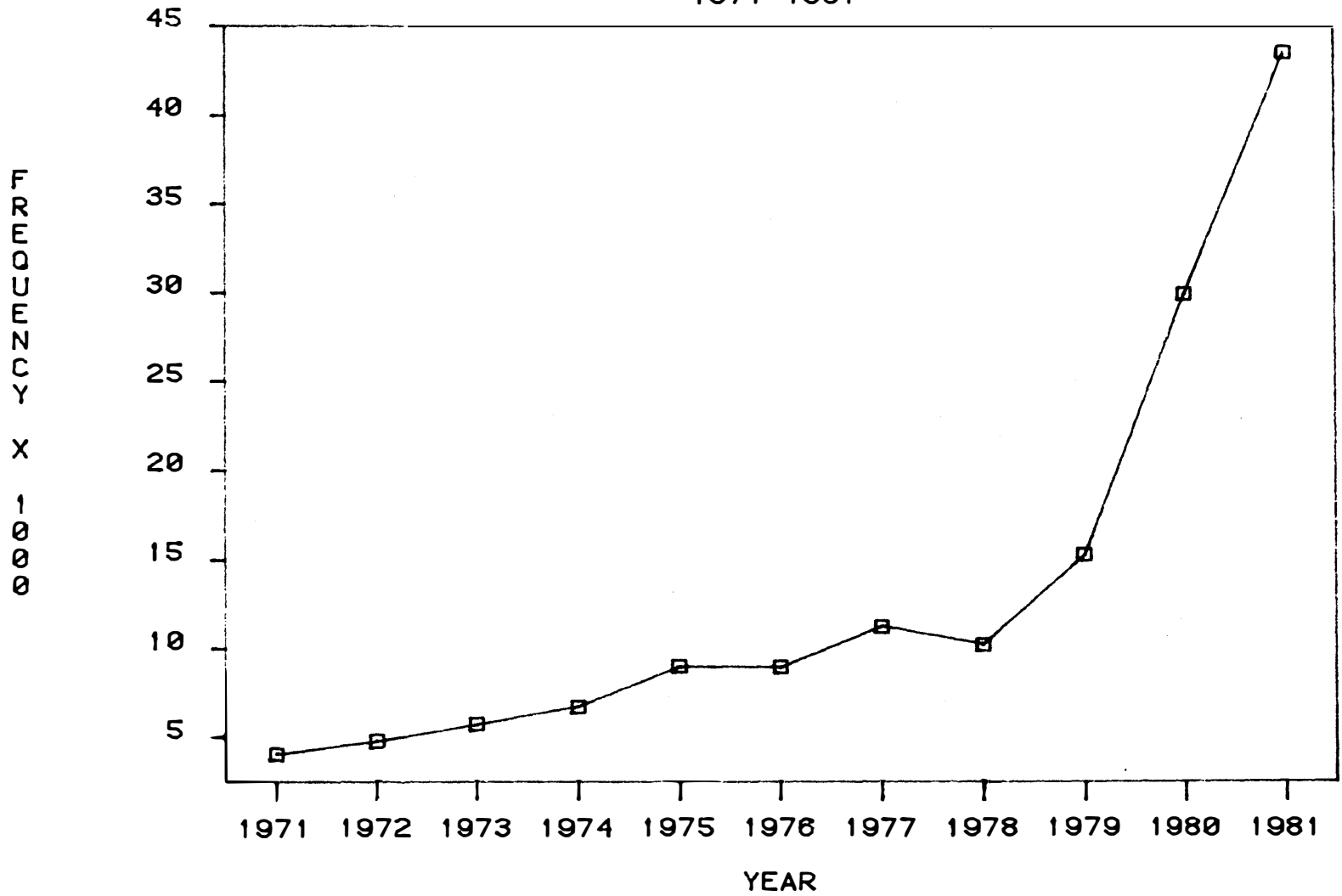
Figure 2

Tables 1 and 2 show that the numbers of individuals in Standard 8 started climbing sharply from 1976. For individuals in Standard 10, the sharp rise occurred, not unexpectedly, two years later in 1978.

Although a much smaller proportion of Blacks reach Standard 8 than Whites, in absolute terms Blacks outnumber Whites in Standard 8. In 1981 there were 136 216 Blacks and 72 835 Whites in this standard. In matric, Whites slightly outnumbered Blacks in 1981: the numbers were 55 667 and 43 556 respectively.

These figures should have made it clear that a definite need exists for cognitive tests for Blacks with Standard 8 and upwards. The number of Blacks with at least 10 years of education entering the job market each year is substantial and in all likelihood will grow considerably in future years. The need for selection and placement instruments will grow accordingly.

In certain cases, tests originally designed for Whites can be used with success on Blacks. For instance, the NIPR regularly uses the spatial 'Blox' test on Blacks with at least Standard 8. On the other hand, the NIPR's tests of general intellectual ability (the various levels of the Mental Alertness) have not always proved to be totally satisfactory for application to Blacks (Hall, 1978). Visser (1978) obtained a mean score of only 16,95 on a sample of 101 first year Fort Hare students on the High Level Mental Alertness Test (which has a maximum possible score of 42). Scores were bunched at the lower end of the scale and the test's KR21 reliability index of 0,72 was lower than is commonly reported for White samples. The reliability of the NIPR's High Level Reading Comprehension test based on the Fort Hare sample was exceedingly low, with a KR20 of only 0,32. The reliability of the Vocabulary test was even lower.

These results suggest the need for a non-verbal test of general intellectual ability suitable for application to Blacks with higher levels of education. In many selection and placement contexts, it is desirable to have an index of a subject's ability to form concepts and

reason abstractly uncontaminated by verbal skills. In tests like the Mental Alertness, the processing of verbal material forms an important component of the test task. Blacks, whose home language is not English or Afrikaans, might have difficulty in understanding what is required of them in certain items. In addition, many items in Mental Alertness - type tests require verbal reasoning. A task of this type is clearly more taxing for a subject whose home language is not the language of the test, for the subject either has to translate the material into his own language before performing logico-verbal manipulations, or perform these manipulations in a second or third language.

A further problem with the use of Mental Alertness - type tests on cultures for which they were not originally intended, is that these tests often have items which are strongly culturally bound. The items frequently have a scholastic flavour about them. As the quality of Black education is generally inferior to that of White, it is not surprising that Blacks have more difficulty doing Mental Alertness - type tasks than Whites. School education is done largely in a verbal medium: concepts are taught through the use of words and logical symbols. Seldom is a pictorial or diagrammatic medium used. The use of a diagrammatic format for a test of reasoning ability or concept formation has the advantage that the scholastic component in the test is reduced as the mode of presentation is not typically scholastic.

Although this scholastic component in a psychological test can be re-duced, it generally cannot be eliminated. Concepts taught in school appear in psychological tests, and individuals who have had less, or poorer quality schooling, can be expected to do more poorly on the tests than those who have had more education or a better quality of it. Nor is it desirable to attempt to eliminate the scholastic component in psychological tests by using concepts which are not taught in Western-type educational institutions. Most tests are used for selection and placement; the ability to recognize and use the sorts of concepts which are taught at school and university probably contribute significantly to the predictive validity of tests . South Africa is

becoming ever more strongly a member of the Western-technological com-
munity of nations; in order to be valid, psychological tests must in-
corporate concepts which are necessary for effective functioning in a
Western-technological society, if these tests are to be used for selection
and placement in the work environment. On the other hand, efforts should
be made to eliminate material from a test which is culture-bound
and also has no predictive validity. Material of this kind is unfair
in the profoundest sense of the word to those individuals who do not
belong to the culture in which the test is sourced, especially if the
same test is being used to select individuals from different cultural
and/or racial backgrounds. It is in an effort (partially) to avoid
this problem than the NIPR adopts the policy of having differential
norm groups; norm groups are fairly precisely described in terms of
race, sex, age, education and work field. This makes it possible to
rate an individual against his peers (as defined in the norm-group
description), but problems arise when an evaluation across groups is
required.

This is a common problem in the practical situation of hiring job
applicants or selecting individuals for training inputs. Ultimately,
the concern is to select people who are capable of handling the intel-
lectual demands placed on them on the job or training course. This is
a problem even when the population to be selected from is homogenous;
if the selection population is heterogeneous, the problem is exacerbated.
There is no perfect solution to this problem. The best compromise is to
determine separate minimum scores for each norm population such that
individuals above that score have a reasonably good chance of being
successful in the activity for which selection is being used and
those below the score are unlikely to be intellectually equipped for
the pursuit in question.

The above procedure involves a large amount of validatory work, ideally
performed specifically for each selection application. In addition, the
problem remains that the test in question might prove to be more valid
for the selection of one type of applicant than another. If this problem
is severe, it will be necessary to use different tests for different
applicant populations.

The High Level Figure Classification Test is a new instrument which in many cases will be a suitable measure to use in the selection of Blacks, Coloureds and Indians with the appropriate educational background for work and training courses which involve a fair amount of conceptual activity. It will give an index of the individual's general intellectual ability uncontaminated by verbal skills. At the job and educational levels where the HL FCT is likely to be used, verbal skills are often important in the successful accomplishment of the tasks in hand. In many cases, therefore, it will be appropriate to apply at least two tests to subjects: the HL FCT and a verbally-based test. (In some instances, yet other tests would be appropriate: for instance a test of number ability or spatial ability if the nature of the work warrants it).

This multitest approach will enable a more accurate assessment of the individual's capacities to be made. When a verbally presented intelligence test is employed, a low score might indicate low verbal skills or poor conceptual ability, or both. The use of an FCT-type test allows conceptual ability to be measured more purely. If verbal, numerical or any other ability is important in the tasks which successful applicants will be required to do, these can be assessed with separate instruments.

As the test uses diagrammatic stimulus material which is likely to be more-or-less equally unfamiliar to all cultural and racial group, it is hoped that the problem of different validities for different types of subjects will be minimized in most applications. The need, however, to establish appropriate selection cut-offs for particular applications, remains.

2.0 BACKGROUND TO THE DEVELOPMENT OF THE TEST

The High Level Figure Classification Test (HL FCT) is very similar to
the standard FCT, both in concept and in item format. The main dif-
ference between the tests is that the HL FCT is designed to be used
on a more highly educated group (individuals with at least Standard 8).

Each item is printed on a separate page; the subject's task is to
divide six diagrams into two groups of three so that the diagrams in
each group share some feature in common. In easier items, the simi-
larities usually are perceptually apparent, whereas in the more dif-
ficult items, the similarities are conceptual. No simple dichotomisa-
tion into perceptual and conceptual can be made; rather perceptual
and conceptual should be seen as opposite ends of a continuum.

Even in the more perceptual items, there is still a substantial
conceptual element. In one of the easier items of the standard level
FCT, for instance, each of the six diagrams incorporates figures which
either decrease in size from left to right, or increase in size from
left to right. The subject has to classify the diagrams into
'increasing' and 'decreasing' categories. Although increasing or
decreasing size is perceptually apparent, the testee has to overcome
distracting influences (such as different shapes and shadings of the
diagrams) in order to identify the relevant concepts.

In most of the items of the HL FCT, the testee receives very little
assistance from perceptually obvious cues. He has to look beyond the
shapes of the diagrams in order to ascertain the underlying concepts.
Let us illustrate this by taking one of the more difficult items
in the HL FCT, one which embodies the notion of 'path-connectedness'.
Two points are path-connected if they can be joined by a line without
passing through any barrier, e.g. :
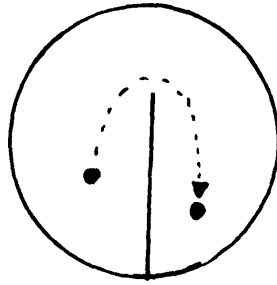
The following points on the other hand are not path-connected:
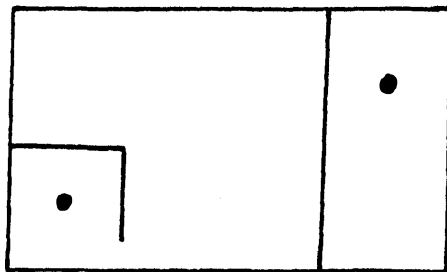
A set of path-connected diagrams can vary enormously in physical appearance and yet all embody the same concept. Similarly, a set of non-path-connected diagrams can differ enormously in appearance.

Up to this point, we have been using a number of terms to describe the task-required in the FCT tests: concept formation, categorization, conceptual functioning etc. Cattell's (1971) term 'fluid intelligence' might also be used to describe the test task. To these we will add one more which we think underlies all these activities: abstraction.

The notion of abstraction has a long history. Aristotle refers to it is his Metaphysics and defines it as the process of disregarding the particular in order to extract what is common. In psychology, the ability to categorize has often been associated with abstracting activity, because by forming a class or category of objects or phenomena one is finding an underlying similarity between these objects while disregarding aspects of dissimilarity.

In Western civilization, the ability to abstract is highly regarded and many psychologists would see it as the basis of intelligent

behaviour. It is hard to imagine the development of Western culture
without simultaneously thinking of abstraction. Spengler (1926)
believes that polyphonic music and the mathematics of functional
relationships are the greatest achievements of modern Western culture
(which he sees dating from about 1 000 A.D. to the present).
Both of these disciplines are highly abstract. Compare for instance the
classical mathematics of Euclidean geometry with calculus. In Euclidean
geometry, the relationships are fixed (pi, for instance) whereas in
calculus, the unknowns in an equation are variable. Variables are highly
abstract concepts. It is impossible to conceive how the scientific,
philosophical and technological successes of Western civilization
could have been achieved without ability to function abstractly. One
must remember, however, that abstractions or concepts are arbitrary.
Cultures form concepts whicn are useful in ordering their work and
social experiences. In categorizing, one is choosing to highlight
certain characteristics while ignoring others.

The converse of abstract thinking is usually called concrete thinking.
Thinking of this kind has often been associated with 'primitive' or
childisn states (e.g. see Lévy-Bruhl, 1923). Some authors
(e.g. Cryns, 1962) see African culture as being highly concretisic
in its thought patterns. Much of this type of commentary can be
accused of cultural chauvinism. Up to the middle of this century
statements about the concreteness of thinking of certain non-Western
peoples were frequent and laden with negative value judgements (see
Werner, 1948 and Carothers, 1953). Authors who made these statements
tended to assume that tests designed for Western subjects were quite
suitable for application to Blacks and would not have any adverse effect
on performance. In many cases, no words existed in subjects' native
language for certain features of the test stimuli which they were
required to handle. Subjects in many cases had absolutely no formal
education and came from cultural environments which were almost
totally untouched by Western thought.

Jahoda (1956) was the first to point to the inadequacies of indiscriminately using Western-designed tests to assess conceptual functioning of non-Western people. He found that a group of non-Westernised illiterate Africans performed significantly worse on a measure of abstract behaviour than a group of relatively more Westernised literate compatriots. Later studies have confirmed this finding (Price- Williams, 1962; Ciborowski and Serpell, 1971, Aborowski and Cole, 1971; Kellagahan, 1968; Okonji, 1971; Irwin and M[c]Laughlin, 1970).

Two other points about the supposed inability of illiterate non-Westernised people to make abstractions will be made here. Firstly it is possible that these people only appear to be concretistic in their thought patterns because they have not been tested on culturally relevant classificatory tasks. Some support for this comes from Okonji (1971) and Price-Williams (1962). Secondly, even if it were possible to establish objectively that certain non-Western cultures are indeed more concretistic, it is a value judgement to say that this type of thinking is unavoidably inferior to abstract thinking. Certainly it is inferior in the context of operating effectively in a Western-technological environment, but it might not be inappropriate in other contexts; in fact it might be beneficial and adaptive in certain cultures. Far from being primitive, the ability to make fine discriminations between objects may be regarded as a demanding cognitive skill. Even in Western society, the previously unquestioned basic tenet of logical and scientific thought that parsimony and generalizability are to be sought at all costs is coming under fire from some critics. These critics would have the Western penchant for classification and reduction tempered with a more holistic orientation, the rationale being that in reducing an object or phenomenon to its constituent parts and classifying it, important uniquenesses of the whole are lost, ignored or distorted. Despite these criticisms, however, it is almost unthinkable that Westerners will completely abandon the logical and scientific cutting tools of classification and reduction, which have served them so well in the building of their materially successful technological society.

The abstraction and classification tasks with which the Figure
Classification Tests confront subjects can be seen as closely related
to tasks which are continually required of people working in a technolo-
gical society. The activity of abstraction/classification involves
both analytic and synthetic elements. The FCT and HL FCT require
these functions in the order: analysis, synthesis, analysis (ASA).

Analysis (1)   The subject first has to scrutinize the stimulus material
                and take note of the precise shapes and components of
                the diagrams.

Synthesis   The subject then has to form hypotheses concerning the
                concepts which are embedded in the item. This is the
                abstracting phase.

Analysis (2)   Finally, the subject has to check whether his hypothesis
                is valid. In particular, he must determine whether there
                are three representatives of each concept to that he
                can achieve the requirement of splitting the six
                diagrams into two groups of three.

In many real-life tasks in Western society, successful functioning is
dependent on the application of an ASA procedure. The individual has
to analyse the complex of phenomena, classify the phenomena into one
or more conceptual classes, then check whether his classification is a
useful and effective one for the situation at hand.

Classification or abstraction can be seen as an information reduction
task (Posner, 1964). It is a "many-to-one" mapping. Many objects,
events or phenomena are mapped into a single conceptual class, as
apples, pears, grapes, etc. can be mapped into the single class of
fruit. Inevitably, much information is lost in this process. However,
if classification is appropriately used, the gains in conceptual economy
outweigh the losses of specific information. Information reduction
classification is essential in Western-technological society for the
following reasons:

1)   Information input is at a very high level in technological society;
      the individual has to classify, simplify and generalize in order
      to avoid being swamped by masses of specific information.

2)   Classification is essential to building up a repertoire of
      appropriate responses to input stimuli. People are confronted
      every day with stimulus complexes differing in various ways
      from situations which they have encountered in the past. The
      ability to take out the essence, to say: "This is an 'abc'
      type situation and hence the best way to handle it is to do
      'xyz'" is essential if one is to be able to put past experience
      of successful behaviour to use in present and future situations.

The HL FCT is considered to be a suitable measure for identifying
individuals who will be required to perform work which is not repeti-
tive, which involves taking decisions without constant reference to a
superior and which requires the incumbent to structure his tasks in
order to manage them effectively.

3.0 STAGES IN THE CONSTRUCTION OF THE HIGH LEVEL FIGURE CLASSIFICATION TEST (HL FCT)

This chapter describes the construction and refinement of the High
Level Figure Classification Test (HL FCT).

The HL FCT was originally constructed with 54 items. This version
was then modified and an improved version of 36 items was produced.
After extensive analysis, the test was further modified and
shortened to 24 items. This 24 item version was classified by the
Test Commission as an 'A' level test in September, 1982.

During the three stages of construction and refinement of the
HL FCT and its subsequent evaluation, ten samples of Asian and
Black subjects with 10 to 12 years of education were tested.
These samples are listed below:

   i) Black employees of a construction company (n=97)
  ii) Asian employees at an insurance company (n=87)
 iii) Black nurses at Baragwanath Hospital (n=451)
  iv) First year Fort Hare students (n=145)
   v) Asian apprentices in the mechanical trades of the sugar
      industry (n=93)
  vi) Black apprentices in the mechanical trades of the sugar
      industry (n=165)
 vii) Black male applicants for technical training (n=263)
viii) First year Mabopane East Technikon students (n=199)
  ix) Black male work-seekers who approached the work orientation
      centre of the Orange-Vaal Adminstration Board (n=126)
   x) Sowetan pupils in Standard 8 doing mathematics and science (n=127)

Detailed analysis and evaluation of the above data resulted in the
final 24 item version of the HL FCT. The main analyses were item
analysis and item response evaluation (see 3.1).

In order to explain the uses of the two programmes, the output statistics of the programmes will be discussed in detail. The data presented for discussion is that of the first sample analyzed, the 97 Black employees of a construction company.

3.1 ANALYSES PERFORMED ON THE 54 ITEM VERSION OF THE HL FCT.

The three versions of the HL FCT were subjected to in-depth analysis. The two programmes used extensively were the NIPR's NP77-A (Coulter, 1973) and NP50 (Maughan-Brown, 1973).

NP77-A, known as the Item Response Evaluation (IRE), was used for evaluating the test as a whole, each individual item and the alternatives within each item. NP50 is an item analysis programme designed for item selection. It was thus used as an aid in choosing the subset of items that made the HL FCT optimal in terms of reliability and distribution form.

Both these programmes have several important features only some of which were relevant to the analysis of the HL FCT.

Only these relevant features will be discussed here. For further information on the programmes, the programme manuals should be consulted (Coulter, 1977; Maughan-Brown, 1973).

3.2 THE NIPR'S NP77-A ITEM RESPONSE EVALUATION PROGRAMME

The NP77-A or the IRE was used for evaluating:

  i)   the test as a whole
 ii)   each individual item
iii)   the alternatives within each item

Each of the above will be considered in turn.

ITEM: 18

No. Attempting = 89   Proportion Attempting = 0,92

Alternative	Omit	1	2	3	4	5	6	7	8	9	10
Prop Endorsing		0,3258	0,0112	0,0112	0,0337	0,0449	0,3034	0,1124	0,0	0,1461	0,0112
Prop Endors Incl Omit	0,0825	0,2990	0,0103	0,0103	0,0309	0,0412	0,2784	0,1031	0,0	0,1340	0,0103
Std. Dev.	0,2751	0,4578	0,1010	0,1010	0,1731	0,1988	0,4482	0,3041	0,0	0,3407	0,1010
Criterion: Key											
Prop Corrected For Guessing		0,2509	0,0125	0,0125	0,0375	0,0499	0,3371	0,1248	0,0	0,1623	0,0125
Mean	26,0000	21,2758	7,0000	30,0000	18,0000	19,2500	23,0370	22,2000	0,0	22,6154	8,0000
Correl. With Crit.	0,1405	-0,1064x	-0,1661	0,0929	-0,0739	-0,0572	0,0880	0,0167	0,0	0,0374	-0,1548
Reliability Index	0,0361	-0,0257	-0,0172	0,0091	-0,0139	-0,0128	0,0304	0,0017	0,0	0,0084	-0,0160
Discrimination Index	0,1419	-0,1070	-0,1684	0,0933	-0,0742	-0,0573	0,0883	0,0167	0,0	0,0375	-0,1567

TABLE 2: Display of IRE Statistics of Item 18 which was dropped from the HL FCT

ITEM: 20

No. Attempting = 94   Proportion Attempting = 0,97

Alternative	Omit	1	2	3	4	5	6	7	8	9	10
Prop Endorsing		0,0532	0,0319	0,0532	0,0319	0,0319	0,6915	0,0213	0,0426	0,0426	0,0
Prop Endors Incl Omit	0,0309	0,0515	0,0309	0,0515	0,0309	0,0309	0,6701	0,0206	0,0412	0,0412	0,0
Std. Dev.	0,1731	0,2211	0,1731	0,2211	0,1731	0,1731	0,4702	0,1421	0,1988	0,1988	0,0
Criterion: Key											
Prop Corrected For Guessing		0,0591	0,0355	0,0591	0,0355	0,0355	0,6572	0,0236	0,0473	0,0473	0,0
Mean	13,0000	11,4000	13,6667	9,8000	8,6667	17,3333	26,0923	17,5000	20,7500	15,2500	0,0
Correl. With Crit.	-0,1714	-0,2664	-0,1578	-0,3091	-0,2600	-0,0828	0,6054x	-0,0645	-0,0150	-0,1456	0,0
Reliability Index	-0,0310	-0,0608	-0,0287	-0,0699	-0,0459	-0,0162	0,2999	-0,0104	-0,0059	-0,0311	0,0
Discrimination Index	-0,1740	-0,2763	-0,1598	-0,3250	-0,2693	-0,0831	0,7606	-0,0646	-0,0150	-0,1471	0,0

TABLE 3: Display of IRE Statistics of Item 20 which was retained for a further version of the HL FCT

Tables 2 and 3
are contained
overleaf.

Please turn over
and open.

3.2.1 TEST CHARACTERISTICS

Table 1 is a display of the test characteristics supplied by the
IRE.

NO. ITEMS	MEAN	STD. DEV.	OMISSION RATIO	K-R 20(1)	K-R 20(2)
54	22,05	9,0754	0,0851	0,8829	0,9002

K-R 20(1) : No response taken as wrong
K-R 20(2) : No response taken as missing information

TABLE 1: Statistics Computed by the IRE of the HL FCT as a Whole

The IRE computes the mean and standard deviation of the test. These
statistics are some of the indicators of the applicability of the
test to the sample tested. The mean provides some information on
the difficulty of the test items, the homogeneity of the sample and
amount of error variance in the test (Coulter, 1977; pp 20 - 21).
For the HL FCT a standard deviation of 9,1 shows that there is a
satisfactory spread of scores across the sample.

The Kuder-Richardson reliability coefficient K-R 20 is an index
of the internal consistency of the test; it takes into account
the correlation of each item with the total test score. In
general, the test constructor is satisfied with coefficients of
0,75 and above. A low coefficient of internal consistency may
indicate that the test is composed of more than one dimension.
The IRE computes two values for the coefficient of internal
consistency. K-R 20(1) is calculated under the assumption that
any items omitted by the respondent could not be done and are
thus wrong. K-R 20(2) assumes that omissions are really unattempted
items and should therefore be taken as missing information. Since
the HL FCT is considered to be a power test with little or no speed
variance, ample time is given to the testee to complete the test.

Thus the first measure of K-R 20 is more appropriate. The HL FCT has high internal consistency (0,88) for the sample tested, taking omitted items as wrong.

Finally, the omission ratio computed by the IRE shows the average proportion of omissions per item. This ratio should always be very low for accurate estimates of the reliability coefficient. For the HL FCT, the omission ratio of 0,08 was considered sufficiently low.

3.2.2 ITEM STATISTICS

The statistics of the 54 item HL FCT considered as a whole seemed satisfactory. The test however was much too long. The next step was thus to examine carefully the statistics of each item and to assess whether or not it should be retained in a shortened version of the test, or discarded. In addition to the statistics descrip-tion above, the IRE provides information on each alternative of every item.

Two items from the original HL FCT have been selected to illustrate the item statistics provided by the IRE. These are displayed in Tables 2 and 3. The relevant parts of these displays have been enclosed in boxes. Table 2 presents statistics on item 18, an item which was dropped from the test; Table 3 shows the statistics for item 20, which was retained.

The IRE computes the number (and proportion) of people who have attempted each item. For items 18 and 20 of the HL FCT, there was a high proportion of endorsements (0,92 and 0,97 respectively). Accurate item statistics can be obtained when item endorsement is high and the sample is large. If a very low number of people endorse an item, it may mean that the item looks difficult to the testee and large number of people do not attempt it for this reason. Such an item should be modified or dropped from the test. If there are items with high proportions of omissions at the end of the test, it may mean that the subjects did not have enough

time to complete the test. In such a case, lengthening the time
limit may be advisable.

The item mean is given by the proportion of respondents endorsing
the correct alternative. In the IRE printout, the column
containing the statistics for the correct alternative is indicated
by an asterisk. The correct alternative has been boxed in Tables
2 and 3. The item mean is an indication of the difficulty level
of the item. The mean for item 18 is 0,3 and for item 20 is 0,67.
Item 20 is therefore a much easier item than item 18. For a
discussion of proportions see page 21.

The item standard deviation is a function of the mean and should,
in most cases, not be small since this will limit the total test
variance. An exception is when the test is to be used to discrimi-
nate at a very high level. In such a case large numbers of items
with very high means are appropriate even at the cost of reducing
item standard deviation and hence test variance. An example of
such a situation is when the tester wishes to select only "best"
people i.e., those at the top extreme of the test score
distribution. On the other hand, a number of items with very low
means would be chosen if the tester wishes to screen out only
those people at the lower tail of the distribution (see 3.2.2 for
a discussion of test score distribution).

For a dichotomous test, the standard deviation of an item is computed
as follows:

$$S_i^2 = P_i(1 - P_i)$$

Where:-
S_i^2 is the variance of item i
and P_i is the mean of item i or the proportion of respondents endorsing
item i correctly

The maximum possible values of the standard deviation and variance of a dichotomous item are 0,5 and 0,25 respectively. These values are attained when half the respondents answer the item correctly.

Both items 18 and 20 have high standard deviations (0,46 and 0,47 respectively).

3.2.3 CHARACTERISTICS OF ALTERNATIVES

Up to this point, the overall test characteristics have been discussed and statistics on the two example items of the HL FCT have been examined. Careful examination of the statistics of the alternatives is also necessary as this might reveal problems with one or more of the alternatives. The IRE provides information on every legitimate way in which the testee can respond.

The HL FCT is a test in which the respondent must group 6 different figures into 2 groups with 3 figures in each. The total number of possible ways of doing this is calculated as follows:

$$^{n}C_r = \frac{n!}{r!(n - r)!} \quad \text{.......... Equation (1)}$$

Where:-
n denotes the total number of different objects
and r denotes the number of objects being selected

For the HL FCT we substitute the values n=6 and r=3, and further divide by 2 since 2 groups of 3 objects each are being selected at once.

Substituting in equation (1) above and dividing by 2, we obtain:

$$\frac{^{6}C_3}{2} = \frac{6!}{3!(6 - 3)!2} = 10$$

Thus there are ten possible ways of responding to any given item
of the HL FCT. The IRE provides statistics relating to each of
these ten individual alternatives.

An important statistic for the test constructor to examine is the
proportion of people in the sample endorsing each alternative.
Since there is only one alternative considered correct for each
item of the HL FCT, there are nine other alternatives all of
which may be thought of as distractors. The test constructor
should be suspicious of any item in which the percentage of
respondents choosing the alternative designated correct is lower
than that for one or more distractors. Such results may be caused
by lack of clarity in the item such as unclear diagrams.

On the other hand, if the distractors of an item lack credibility,
few or none of the respondents will endorse them and the probability
of guessing the correct answer may be increased. Transparently
incorrect alternatives should thus be reviewed or the item should
be replaced.

Thus a good item is one where the proportion of respondents endorsing
any of the distractors is neither higher than the proportion
endorsing the correct alternative, nor near zero.

The IRE provides three sets of proportions of endorsements on the
alternatives of each item.

  i) proportions excluding omissions
 ii) proportions including omissions
iii) proportions corrected for guessing

The first set of proportions is most appropriate when omitted
items are taken as missing information rather than as incorrect
responses. As discussed earlier, any items omitted from the
HL FCT were considered to be wrong rather than unattempted.
Consequently, the set of proportions including omissions is
appropriate for the test.

The IRE provides an estimate of the proportion of testees who really know the correct answer. This estimate attempts to exclude those individuals who guessed the correct answer. This third set of proportions is computed under the assumption that testees who do not know the answer guess randomly (see Coulter, 1977; pp 27 - 29). Thus the corrected proportion endorsing the correct alternative will be deflated while the corrected proportions endorsing the distractors will be slightly inflated.

Let us return to Table 2 and examine the proportion of the respondents (including omissions) endorsing the alternatives of item 18. Only 0,3 or 30% of the respondents endorsed the alternative designated correct. In itself, there is nothing alarming about this figure; however on further examination, it can be noticed that approximately 0,28 or 28% of the respondents endorsed the 6th alternative. After correction for guessing an even worse picture emerges: the 6th alternative was endorsed more often than the 'correct' alternative, and there was little endorsement of 6 of the other 8 distractors. In such a case there may be ambiguity or there may be an alternative correct response. A characteristic of tests like the HL FCT which involve inductive tasks is that there is no unique correct answer. The test constructor goes out of his way to build a salient concept into each item. If he does not succeed in his aim, respondents might find another 'correct' answer.

The distribution of responses across the alternatives of item 20 (Table 3) looks better. While approximately 0,66 or 66% of the respondents endorsed the correct alternative, more or less equal proportions of the respondents endorsed the distractors. This shows the absence of both ambiguity and alternative correct solutions to the item.

Once the test constructor has looked at the distribution of responses across the alternatives, he should see whether the 'brighter' people in the sample are endorsing the correct alternative. By inspecting the average score on the whole test of those respondents endorsing each alternative, it can be seen which alternative is

being endorsed by the 'brighter' respondents. This average test
total is given by the IRE as the mean for each alternative.

For a good item, the correct alternative should be endorsed by the
brighter people and have a mean (or average test total) which is
higher than that of any of the distractors.

If the 'bright' people or those who do well on the test endorse
the correct alternative, this alternative will correlate highly
with the test total and the correlation between the distractors
and the total will be negative. Any item in which one of the
distractors correlates more highly with the test total than does
the correct alternative should be carefully inspected for ambiguity
or an alternative solution apparent only to high scoring
respondents. Items which are excessively easy or difficult may
correlate near zero with the test total. Such items provide little
discrimination between people and should be revised or discarded.

Examining item 18 (Table 2), one can immediately see the reason
for dropping this item from the test. The mean for the alternative
designated correct (21,28) is lower than that for several of the
distractors, and the correlation for the 'correct' alternative with
the total (or criterion) is negative (-0,11). This indicates that
the people who scored well on the test as a whole did not endorse
that alternative. Moreover, the correlations of the distractors
with the total are often positive. These statistics show that
there is something seriously wrong with the item.

On the other hand, the patterns of means and correlations of the
alternatives of item 20 are quite different. The correlations of
the correct alternative with the total is high (0,61) and those
correlations between the distractors and the total are all
negative. Such an item is valuable to include in the test.

Two further sets of statistics on alternatives are provided by the
IRE for the HL FCT. These are the reliability index and the
discrimination index. These indices however, were not the main

criteria considered in the evaluation of the items of the HL FCT
and hence will be discussed only briefly.

For the correct alternative, the reliability index gives an
indication of the contribution of the item to the total test
reliability. In order to obtain good discrimination between
respondents, it may be necessary to include items which have
relatively low standard deviations and lower reliability indices
as a consequence. If these items have high correlations with
the total, they should still be included in the test in spite
of their relatively lower reliability index (Coulter, 1977).

If one examines the set of reliability indices for item 18, it
can be seen that all these values are near zero. For the
correct alternative, the reliability is near zero (-0,03) and
lower than that of some of the distractors. The picture is once
again different for item 20. The reliability index for the correct
alternative is fairly high (0,3) and that for each of the distractors
negative.

The discrimination index for the correct alternative is an estimate
of the maximum discriminating power of the item. Loosely speaking,
item discriminating power is an index of the effectiveness of an
item in discriminating between "good" and "poor" subjects. The
maximum discrimination takes place at a point that is a function
of the proportion of respondents endorsing the item (Lord and
Novick, 1968).

Returning once again to the statistics of items 18 and 20, it can
be seen that the discrimination indices are consonant with other
indices of performance for these items.

The discrimination index for the correct alternative of item 18
is negative (-0,11) and lower than that of most of the distractors;
item 20 has high discrimination index (0,76).

In the above disccusion, only the statistics used in the construction and the refinement of the HL FCT have been reviewed. For a complete and more detailed discussion of the capabilities of the IRE see Coulter (1973).

The other programme used extensively in the refinement of the HL FCT is the NIPR's NP50 item analysis programme. The following section describes this programme as used for the HL FCT.

3.3   THE NIPR'S NP50 ITEM ANALYSIS PROGRAMME

Unlike the IRE which is intended for evaluating items individually and the test as a whole, the item analysis programme or NP50 is designed for item selection. It was thus used as an aid in choosing the subset of items that makes the HL FCT optimal in terms of reliability and distribution form. The programme can iteratively maximize the reliability of the test, dropping those items which fall below a specified criterion level until maximum reliability has been obtained. Once again, only the programme options used for the HL FCT will be described here. Further information may be obtained from the programme manual (Maughan-Brown, 1973).

For the original set of items as well as for each successive iteration, NP50 supplies the following:
  i)   test score statistics
 ii)   distribution form
iii)   item-test correlations
 iv)   means and standard deviations of item statistics
  v)   reliability coefficients

3.3.1 TEST CHARACTERISTICS

In the case of a dichotomous (right-wrong) test such as the HL FCT, the NP50 programme required a key which indicates the correct response for each item considered. NP50 computes total test scores by counting the number of correct responses for each subject. It then calculates the mean and standard deviation of the test scores, giving the 95% confidence limits for both. Table 4 below is a display of these statistics for the HL FCT. It may be observed that the mean and standard deviation correspond to those in Table 1 since both programmes compute these statistics. As in the previous sections, all statistics quoted in this chapter are based on the sample of 97 construction workers.

N	MEAN	95% LIMITS OF MEAN	SD	95% LIMITS OF SD
97	22,052	20,222 - 23,881	9,075	7,953 - 10,569

TABLE 4: Test Score Characteristics of the HL FCT computed by the Item Analysis Programme (NP50)

3.3.2 DISTRIBUTION FORM

The programme supplies information on the form of the distribution of scores. A histogram of the scores is plotted and the maximum and minimum test scores and the score range printed. Table 5 below is a display of the distribution form of the scores as computed by the NP50 programme. If the maximum possible score is greater than 15, the scores are compressed into a 15 unit histogram labelled from 0 to 14.

DISTRIBUTION OF TEST SCORES		
0	0,667	
1	4,333	XXX
2	8,000	XXXXX
3	7,667	XXXX
4	9,333	XXXXXX
5	12,000	XXXXXXXX
6	19,667	XXXXXXXXXXXXX
7	10,667	XXXXXXX
8	10,667	XXXXXXX
9	10,000	XXXXXX
10	2,667	XX
11	1,333	X
12	0,0	
13	0,0	
14	0,0	
Maximum Test Score = 40		
Minimum Test Score = 3		
Range = 37		

TABLE 5: Distribution of Scores on the HL FCT computed by the Item

Analysis Programme.

The shape or form of the distribution is very important. The
distribution shape is determined by the difficulty level of the
items, the covariances among items, and the number of items
(Nunally, 1978).

A distribution tends to be skewed whenever the test mean approaches
scores on the upper or lower extreme of the distribution. Distri-
butions which are positively skewed or skewed toward the upper
end of the scale are often produced by tests with items that
are relatively difficult. In such cases the majority of test scores
are fairly low and thus the test mean is fairly low. Conversely,
negatively skewed distributions are often produced when the respondents
find the items relatively easy resulting in fairly high test total
mean.

The test constructor may wish to construct a test with a
distribution which approaches the symmetric bell-shaped curve
of the normal distribution. However, in many cases a flattened
normal or platykurtic distribution is desirable as tests with
this type of distribution generally discriminate fairly well
across the entire score range. Skewed distributions are
nevertheless often useful for stringent selections of people
at the extremes of the distribution.

The distribution of HL FCT scores in Table 5 is positively
skewed. The mean score (22,05) is fairly low for a test of 54
items and no respondent scored higher than 40. This version of
the test is thus fairly difficult for the sample concerned.

3.3.3 ITEM STATISTICS

The programme computes several item statistics which are useful
to the test constructor in item selection and evaluation.
These statistics are the same as those computed for the correct
alternatives of each item by the IRE and hence will be
discussed only briefly. In instances where the detail of the
IRE analysis is not required, the summary form provided by the
NP50 programme is sufficient.

The item statistics computed by NP50 include the following:

PJ the proportion of respondents who answered the item incorrectly

PW the proportion of respondents who answered the item correctly. Therefore PW = 1 - PJ

SJ2 the variance of the test totals of those respondents who correctly answered the item

SJ the standard deviation of the test totals of those respondents who answered the item incorrectly. SJ = SJ2

XBARJ the mean of the test totals of the respondents who correctly answered the item

RX the item-test point biserial correlation i.e., the correlation computed for all respondents between each person's score on the item and his test total. The point-biserial correlation is computed whenever there is one dichotomous variable (such as scores on a right-wrong item) and one continuous variable (such as test totals)

RXSJ Gulliksen's Index is the item-test point biserial correlation multiplied by the standard deviation of the item

ITEM-TEST CORRELATIONS

No. of Items in the Test Score = 54

ITEM	NJ	PJ	PW	SJ2	SJ	XBARJ	RXSJ	RX
1	41	0,423	0,577	0,244	0,494	25,854	0,178	0,360
2	41	0,423	0,577	0,244	0,494	28,000	0,278	0,564
3	64	0,660	0,340	0,224	0,474	25,906	0,282	0,595
4	50	0,515	0,485	0,250	0,500	27,180	0,293	0,586
5	48	0,495	0,505	0,250	0,500	27,604	0,304	0,609
6	57	0,588	0,412	0,242	0,492	25,860	0,248	0,504
7	64	0,660	0,340	0,224	0,474	25,656	0,263	0,556
8	74	0,763	0,237	0,181	0,425	25,203	0,266	0,626
9	80	0,825	0,175	0,145	0,380	24,563	0,229	0,603
10	44	0,454	0,546	0,248	0,498	27,886	0,293	0,589
11	74	0,763	0,237	0,181	0,425	25,446	0,287	0,674
12	38	0,392	0,608	0,238	0,488	24,447	0,104	0,213
13	38	0,392	0,608	0,238	0,488	25,395	0,145	0,297
14	52	0,536	0,464	0,249	0,499	26,058	0,238	0,477
15	75	0,773	0,277	0,175	0,419	23,653	0,137	0,328
16	47	0,485	0,515	0,250	0,500	24,319	0,122	0,244
17	33	0,340	0,660	0,224	0,474	27,303	0,198	0,418

TABLE 6: Item-test Correlations Computed by the Item Analysis Programme (NP50) on the HL FCT Data

ITEM-TEST CORRELATIONS

ITEM	NJ	PJ	PW	SJ2	SJ	XBARJ	RXSJ	RX
18	29	0,299	0,701	0,210	0,458	21,276	-0,026	-0,056
19	32	0,330	0,670	0,221	0,470	27,469	0,198	0,421
20	65	0,670	0,330	0,221	0,470	26,092	0,300	0,638
21	38	0,392	0,608	0,238	0,488	27,263	0,226	0,463
22	55	0,567	0,433	0,246	0,495	25,855	0,239	0,482
23	29	0,299	0,701	0,210	0,458	29,000	0,230	0,503
24	35	0,361	0,639	0,231	0,480	25,543	0,140	0,291
25	67	0,691	0,309	0,214	0,462	25,045	0,229	0,495
26	44	0,454	0,546	0,248	0,498	25,750	0,186	0,373
27	27	0,278	0,722	0,201	0,448	22,519	0,014	0,032
28	56	0,577	0,423	0,244	0,494	25,464	0,218	0,442
29	61	0,629	0,371	0,233	0,483	25,410	0,234	0,484
30	46	0,474	0,526	0,249	0,499	28,043	0,315	0,630
31	17	0,175	0,825	0,145	0,380	25,353	0,064	0,169
32	19	0,196	0,804	0,158	0,397	28,684	0,144	0,363
33	24	0,247	0,753	0,186	0,432	28,167	0,168	0,388
34	59	0,608	0,392	0,238	0,488	25,000	0,199	0,407
35	47	0,485	0,515	0,250	0,500	24,277	0,119	0,239
36	27	0,278	0,722	0,201	0,448	28,259	0,191	0,427

TABLE 6 Continued: Item-test Correlations Computed by the Item Analysis Programme (NP50) on the HL FCT Data

ITEM-TEST CORRELATIONS

ITEM	NJ	PJ	PW	SJ2	SJ	XBARJ	RXSJ	RX
37	33	0,340	0,660	0,224	0,474	25,879	0,144	0,304
38	37	0,381	0,619	0,236	0,486	26,649	0,194	0,400
39	35	0,361	0,639	0,231	0,480	26,057	0,160	0,333
40	55	0,567	0,433	0,246	0,495	21,873	-0,011	-0,023
41	46	0,474	0,526	0,249	0,499	26,239	0,220	0,440
42	28	0,289	0,711	0,205	0,453	27,036	0,159	0,352
43	4	0,041	0,959	0,040	0,199	21,750	-0,001	-0,007
44	20	0,206	0,794	0,164	0,405	27,450	0,123	0,305
45	4	0,041	0,959	0,040	0,199	25,500	0,016	0,079
46	13	0,134	0,866	0,116	0,341	27,154	0,076	0,222
47	47	0,485	0,515	0,250	0,500	25,468	0,183	0,367
48	42	0,433	0,567	0,246	0,495	25,048	0,144	0,290
49	8	0,082	0,918	0,076	0,275	25,125	0,028	0,102
50	13	0,134	0,866	0,116	0,341	27,231	0,077	0,226
51	16	0,165	0,835	0,138	0,371	27,688	0,103	0,277
52	5	0,052	0,948	0,049	0,221	27,000	0,028	0,128
53	20	0,206	0,794	0,164	0,405	26,000	0,090	0,223
54	16	0,165	0,835	0,138	0,371	24,313	0,041	0,111

TABLE 6 Continued: Item-test Correlations Computed by the Item Analysis Programme (NP50) on the HL FCT Data

3.3.4 MEANS AND STANDARD DEVIATIONS OF ITEM STATISTICS

NP50 computes the overall mean and standard deviation of the sta-
tistics provided for the individual items included in the test.
Table 7 shows an extract of these statistics.

MEANS AND STANDARD DEVIATIONS OF ITEM STATISTICS			
	PJ	RXSJ	RX
MEAN	0,408	0,167	0,362
SD	0,202	0,092	0,185

TABLE 7:   Means and Standard Deviations of Item Statistics

of HL FCT Data Computed by NP50

On examining the figures in Table 7, it can be seen that the
difficulty of a hypothetical average item is 0,408, i.e.,
40,89 of the respondents would have answered such an item correctly.
Since less than half of the respondents answered correctly, the
testees found the test fairly difficult resulting in a positively
skewed distribution. This was shown in Table 5.

Turning now to the statistic RXSJ or Gulliksen's index, it may
be seen that the mean is 0,167 over all items. Gulliksen's
index is an extremely useful criterion for deciding whether to
retain or to drop items when refining a test. This index is
often preferred to the straightforward item-test point
biserial correlation (RX) since it takes into account the standard
deviation of the item.

3.3.5 <u>RELIABILITY COEFFICIENTS</u>

The programme calculates four coefficients of reliability. Of
these, the Kuder-Richardson 20 (K-R 20) coefficient is by far
the most popular index of internal consistency. The less
popular K-R 14 estimate of reliability assumes equal item inter-
correlations (see Kuder and Richardson, 1937; Horst 1953b).
The other two reliability indices provided, the Horst and Loevinger
(see Horst 1953 a), are infrequently used.

Table 8 below is a display of the reliability coefficients
supplied by the programme.

RELIABILITY COEFFICIENTS	
Kuder-Richardson 20	0,883
Kuder-Richardson 14	0,883
Loevinger	0,197
Horst 21	0,893

TABLE 8: <u>Reliability Coefficients of the HL FCT</u>

<u>Computed by NP50</u>

The K-R 20 index indicates that the HL FCT has a reliability
comfortably in excess of the "rule of thumb" lowest
acceptable minimum of 0,75.

3.3.6 <u>THE ITERATIVE FACILITY OF NP50</u>

Ideally, the test constructor aims to produce a homogeneous set
of items in order to create a unidimensional test. To achieve
a test of such internal consistency, the item pool should be
successively refined. In the case where the bulk of the items
measure a single dimension, the test can be made increasingly
unidimensional by successively dropping those items which do
not meet certain criteria. The test and item statistics should

then be recomputed on the reduced pool of items to evaluate whether the test has been improved. This process then continues, with weaker items being dropped from the test on each successive occasion until the test constructor is satisfied that an optimal subset of items in terms of internal consistency, distribution form and number has been selected from the original set of items.

NP50 can be used to achieve the above-mentioned aims. The programme can be used iteratively, dropping items which fall below a specified item-test correlation level. The programme user can choose such criteria as RX or RXSJ to use in item selection. He specifies too, the value of the critical index and the step size by which this index is increased on each iteration. The iteration procedure stops either when the maximum number of iterations specified by the user has been reached or when the reliability (K-R 20) has decreased on two successive iterations.

Tables 9 a - e show the seventh iteration performed on the test items. By this time the programme has dropped 18 of the 54 items. At each stage of the iteration procedure, revised test score statistics, distribution of test scores, item-test statistics, means and standard deviations of item statistics and reliability coefficients are computed.

N	MEAN	95% LIMITS OF MEAN	SD	95% LIMITS OF SD
97	17,711	16,078 - 19,348	8,105	7,103 - 9,439

TABLE 9a: Test Score Characteristics of the HL FCT computed by NP50

on the Seventh Iteration

DISTRIBUTION OF TEST SCORES		
0	1,933	X
1	4,867	XXXX
2	5,000	XXXX
3	6,933	XXXXXX
4	7,267	XXXXX
5	5,200	XXXX
6	8,400	XXXXXX
7	15,800	XXXXXXXXXXXXXX
8	8,000	XXXXXX
9	10,267	XXXXXXXX
10	9,733	XXXXXXXX
11	6,000	XXXX
12	4,667	XXXX
13	2,467	XX
14	0,467	

Maximum Test Score = 34

Minimum Test Score = 1

Range = 33

TABLE 9b: Distribution of Scores on the HL FCT computed by NP50 on the

Seventh Iteration

MEANS AND STANDARD DEVIATIONS OF ITEM STATISTICS			
	PJ	RXSJ	RX
MEAN	0,492	0,224	0,475
SD	0,164	0,058	0,121

TABLE 9d: Means and Standard Deviations of Item Statistics

of HL FCT Data Computed by NP50 on the Seventh Iteration

ITEM-TEST CORRELATIONS

No. of Items in the Test Score = 36
(Items that are blank or marked with - are excluded from the test score)

ITEM	PJ	PW	XBARJ	RXSJ	RX
1	0,432	0,577	21,024	0,174	0,352
2	0,432	0,577	22,805	0,267	0,541
3	0,660	0,340	21,281	0,292	0,617
4	0,515	0,485	22,680	0,318	0,636
5	0,495	0,505	22,917	0,319	0,639
6	0,588	0,412	21,386	0,268	0,544
7	0,660	0,340	20,844	0,256	0,541
8	0,763	0,237	20,581	0,272	0,638
9	0,825	0,175	20,087	0,243	0,639
10	0,454	0,546	23,250	0,312	0,626
11	0,763	0,237	20,865	0,298	0,702
-12	0,392	0,608	19,132	0,069	0,141
13	0,392	0,608	20,684	0,144	0,296
14	0,536	0,464	21,731	0,267	0,536
15	0,773	0,227	19,213	0,144	0,344
-16	0,485	0,515	19,532	0,109	0,219
17	0,340	0,660	22,606	0,207	0,436
-18	0,299	0,701	16,138	-0,058	-0,127

TABLE 9c: Item-test Correlations Computed by the NP50 on the HL FCT Data on the Seventh Iteration

ITEM-TEST CORRELATIONS

ITEM	PJ	PW	XBARJ	RXSJ	RX
19	0,330	0,670	22,750	0,206	0,438
20	0,670	0,330	21,415	0,308	0,655
21	0,392	0,608	22,553	0,235	0,482
22	0,567	0,433	21,218	0,247	0,498
23	0,299	0,701	24,000	0,233	0,509
24	0,361	0,639	20,914	0,143	0,298
25	0,691	0,309	20,522	0,241	0,521
26	0,454	0,546	20,636	0,165	0,331
-27	0,278	0,722	17,296	-0,014	-0,032
28	0,577	0,423	20,768	0,219	0,443
29	0,629	0,371	20,852	0,245	0,507
30	0,474	0,526	22,848	0,302	0,605
-31	0,175	0,825	19,941	0,048	0,127
32	0,196	0,804	23,684	0,145	0,366
33	0,247	0,753	22,833	0,157	0,364
34	0,608	0,392	20,220	0,189	0,388
-35	0,485	0,515	19,468	0,106	0,211
36	0,278	0,722	23,370	0,195	0,436
37	0,340	0,660	21,485	0,159	0,336
38	0,381	0,619	21,541	0,181	0,373
39	0,361	0,639	21,514	0,170	0,354

TABLE 9c: Item-test Correlations Computed by the NP50 on the HL FCT Data on the Seventh Iteration

ITEM-TEST CORRELATIONS

ITEM	PJ	PW	XBARJ	RXSJ	RX
-40	0,567	0,433	17,418	-0,021	-0,042
41	0,474	0,526	21,522	0,224	0,449
42	0,289	0,711	21,857	0,148	0,328
-43	0,041	0,959	16,000	-0,009	-0,044
-44	0,206	0,794	21,950	0,108	0,268
-45	0,041	0,959	20,000	0,012	0,059
-46	0,134	0,866	21,231	0,058	0,172
47	0,485	0,515	20,511	0,168	0,337
-48	0,433	0,567	19,810	0,113	0,227
-49	0,082	0,918	20,125	0,025	0,090
-50	0,134	0,866	21,385	0,061	0,179
-51	0,165	0,835	21,688	0,081	0,219
-52	0,052	0,948	20,200	0,016	0,072
-53	0,206	0,794	20,700	0,076	0,189
-54	0,165	0,835	18,750	0,021	0,057

TABLE 9c: Item-test Correlations Computed by the NP50 on the HL FCT Data on the Seventh Iteration

RELIABILITY COEFFICIENTS	
Kuder-Richardson 20	0,901
Kuder-Richardson 14	0,901
Loevinger	0,291
Horst 21	0,912

TABLE 9e: Reliability Coefficients of the HL FCT Computed by NP50 on the Seventh Iteration

From Tables 9a,b,c,d and e it may be seen that the programme has dropped 18 of 54 items by the seventh iteration. It has recalculated all statistics based on the reduced item pool. The new statistics are much improved for the following reasons:

x The test mean is nearer the centre of the distribution and the standard deviation is satisfactory;

x The test score distribution is approaching the normal distribution and is no longer distinctly positively skewed;

x The maximum test score obtained is closer to the maximum possible score than was the case for the 54 item version;

x The average difficulty value of an item has increased from 0,408 to 0,492 indicating that about half of the respondents would respond correctly to the "hypothetical average item";

x The reliability of the test is extremely high indicating a test with high internal consistency.

Thus, by the process of iterative item analysis, the optimal subset of original items can be selected.

The item response evaluation and item analysis described above are only two of the many analysis which were carried out in the construction of the 24 item version HL FCT. There followed detailed analysis of the statistics obtained from all of the ten samples tested. This process finally lead to a test of 24

items which had highly acceptable characteristics for the population
groups for which it is intended.

As was mentioned previously the HL FCT went through three major
stages in its development. In this process the number of items
was reduced from 54 to 36 and finally to 24. Table 10 summarises
this development.

ANALYSIS NO.	SAMPLE TESTED	ANALYSIS
1	97 Black Construction Workers	Iterative Item Analysis
2	97 Black Construction Workers	Item Response Evaluation
3	87 Asian Insurance Workers	Iterative Item Analysis
4	87 Asian Insurance Workers	Item Response Evaluation
5	184 Black and Asian Workers (above samples combined)	Item Response Evaluation on: i) 54 items ii) 35 items iii) 38 items
6	184 Black and Asian Workers	Item Analysis on 36 Selected Items
7	451 Black Nurses at Baragwanath Hospital	Item Analysis
8	451 Black Nurses at Baragwanath Hospital	Item Response Evaluation
9	145 First-year Fort Hare Students	Item Analysis
10	145 First-year Fort Hare Students	Item Response Evaluation
11	93 Asian Sugar Workers	Item Analysis
12	165 Black Sugar Workers	Item Analysis
13	263 Black Male Applicants for Technical Training	Item Analysis
14	199 First-year Mabopani East Technikon Students	Basic Statistics (means, standard deviation, reliability, distribution form)
15	126 Black Work-seekers	Basic Statistics
16	127 Sowetan Pupils	Basic Statistics

TABLE 10: Analyses carried out on the data collected from 10 samples which lead to the final version HL FCT

4.0   EVALUATION AND VALIDATION OF THE HL FCT   (24 ITEM VERSION)

The 24 item version HL FCT was administered to six samples. The first section of this chapter describes the basic statistics of the HL FCT on these samples. The second section presents detailed descriptions of four samples used for the validation of the test. Thereafter, the validation of the HL FCT is discussed for each of these samples. The final section compares the HL FCT and the Mental Alertness Test and argues for the inclusion of both tests in many prediction procedures.

4.1   MEANS, STANDARD DEVIATIONS AND OTHER BASIC STATISTICS OF THE HL FCT

The 24 item version of the HL FCT was administered to six samples (samples v to x listed on page 14). On the basis of the data collected from these samples, the test was critically evaluated. Special attention was paid to the test mean, variance, reliability and distribution form in each of the six samples.

Histograms of the distribution of scores for each of the samples tested are found in Appendix (i).

The sample size, mean, variance, standard deviation and coefficient of internal consistency (K-R 21) for the six samples are presented in Table 11.

SAMPLE	v ASIAN SUGAR WORKERS	vi BLACK SUGAR WORKERS	vii BLACK MALE APPLICANTS FOR TECHNICAL POSITIONS	viii FIRST YEAR MABOPANE EAST TECHNIKON STUDENTS	ix BLACK MALE WORK- SEEKERS AT O.V. ADMIN. BOARD	x SOWETAN MALE PUPILS
n	93	165	263	199	126	127
mean	15,04	13,85	12,65	14,83	10,27	16,91
variance	22,32	32,25	29,67	26,93	30,1	17,13
standard deviation	4,27	5,67	5,45	5,19	5,49	4,14
reliability K-R 21	0,781	0,784	0,833	0,824	0,84	0,74

TABLE 11:   Sample size, mean, variance, standard deviation and reliability statistics

for the HL FCT on each of the six samples tested.

Several observations based on these statistics follow:

Apart from the score distirbution of the Black work-seekers (sample
ix) all distributions are similar in form. There appears to be a
slight positive skewness in all samples except sample ix. Distributions
of this form can be most useful in industry where it is required to
select out those who would not succeed on the job. The similarity
of the distributions suggests that the test may be used for Asian
and Black race groups. Further testing may reveal the applicability
of the test to other race groups. The relatively high frequency of
very low scores in the distribution of the Black work-seekers
suggests that there may be a number of respondents with less than
the required minimum of ten years of schooling or that the greater
proportion of older subjects in this sample was responsible for
the low scores.

The mean score of 12,65 for the applicant group (sample vii) is in
the middle of the distribution for a test consisting of 24 items.
The relatively low mean of 10,27 for the Black work-seekers is a
result of the large number of very low test scores. The mean
scores for all the other groups are higher. For the matriculated
Asian and Black sugar workers, the first-year Technikon students
and the Sowetan matric students, the relatively high mean scores
(15,04; 13,85; 14,83; 16,91) show that the HL FCT scores generally
increase with increasing years of education. The high mean score
for the Sowetan pupils is not unexpected since this was a fairly
highly selected sample of pupils studying matric mathematics
and science. When one considers the Black samples only, it is
interesting to observe that the mean scores for the HL FCT are
highest in the Technikon and Sowetan students. This may be because
people in an educational contest are more accustomed to tests and
examinations than industrial workers and therefore score better
in a testing situation. Another factor favoring the students over
the industrial workers is that they are younger. However, the
HL FCT is still suitable for testing at the upper education range
since only a small percentage of each of the sample scores above
20.

The coefficients of internal consistency (K-R 21) quoted above
are considered satisfactory. The relatively lower coefficient of
consistency for the Sowetan group (0,74) is the result of the
highly homogeneous nature of the sample. One should bear in mind
that this sample has the lowest variance of all (17,13). It should
be remembered too that the K-R 21 coefficient is actually an
underestimate of the K-R 20 coefficient of reliability.

4.2 DESCRIPTION OF THE SAMPLES USED FOR VALIDATION PURPOSES

For this section of the report it is necessary to give a detailed
description of each of the samples which provided validation data
for the HL FCT. Validation data was available for four of the six
samples.

4.2.1   SAMPLES v AND vi

Two samples of apprentices in the mechanical trades of the sugar
industry were tested on the HL FCT. All these subjects were
matriculants who had passed matric mathematics and science.

Sample v consisted of 93 Asians with age range 16 - 26 years and
average age = 19,9 years.

Sample vi consisted of 165 Blacks with age range 17 - 33 years
and average age = 22,1 years.

The B/2/2 version of the Mental Alertness (MA), Mechanical
Comprehension Test (A/3/1), Blox, Technical Reading Comprehension
from the General Science Test (TRC) and Gottschaldt Figures
Test (GFT) were administered to the Asian sample and to 72 of the
Black sample. In addition, the matric mathematics and science
marks were obtained to serve as additional criterion data for
validation of the HL FCT.

4.2.2   SAMPLE vii

In 1980 a sample was collected of 263 Black males who applied for
training as technicians for electronic and electromechanical
equipment (adding machines, copiers, etc.). The age range
was 17 - 29 years and education range 8 - 10 years.

These applicants were given a battery of tests comprising the
HL FCT and other NIPR tests - the Mental Alertness Intermediate
Level (IMA), A/3/1, Blox, TRC, as well as two spatial tests - the
'F' and the 'H' (see Taylor, 1980 for descriptions of the 'F' and
'H' tests). On the basis of these test results and interviews,
a sample of 66 Blacks was selected. After one year's training,
criterion data were collected in the form of end-of-year tests
and examination results.

4.2.3 SAMPLE viii

In February 1982, 199 first year students of the Mabopani East Technikon were tested to assess the feasibility of using NIPR tests for selection of students for the Science Department of the Technikon.

Tests administered to this sample were the HL FCT, IMA, BLOX, TRC as well as two new experimental NIPR tests - the Estimation Test (ESTIM) and the Standard Level Arithmetic Reasoning Test (SL ART).

4.3 VALIDATION OF THE HL FCT

The valication of the HL FCT will be discussed separately for the Asian and Black sugar workers (samples v and vi), the Black male applicants for technical positions (sample vii), and the first year Mabopani East Technikon students (sample viii).

4.3.1 SAMPLES v AND vi

The correlations between the HL FCT and the mathematics and science criterion scores for the Asian and Black Sugar Workers are:

Sample / Criterion	ASIANS (n = 93)	BLACK (n = 72)
mathematics	$0,29^{x}$	$0,32^{x}$
science	$0,08$	$0,40^{x}$

TABLE 12:   HL FCT Correlations with Matric Mathematics and Science Scores for Asian Sugar Workers (n=93) and Black Sugar Workers (n=72)

^{x} indicates correlations significantly different from 0 ($p \leqslant 0,05$).

The correlations between the HL FCT and the matric mathematics
marks for both samples are significant.

The correlations between the science mark and the HL FCT is
significant only for the Black sample. Upon examining Table 13
it can be seen that no test correlates significantly with the
science mark. It is possible that something went amiss either
in the collection of the data or in the science examination
itself.

The complete intercorrelation matrix for the Asian Sugar Workers
is set out in Table 13.

	MATHS	SCIENCE	MA	HL FCT	A/3/1	BLOX	TRC	GFT
MATHS								
SCIENCE	0,58 *							
MA	0,31 *	0,15						
HL FCT	0,29 *	0,08	0,61 *					
A/3/1	0,16	0,18	0,42 *	0,38 *				
BLOX	0,16	0,16	0,41 *	0,55 *	0,29 *			
TRC	0,22 *	0,12	0,59 *	0,28 *	0,20	0,22 *		
GFT	0,12	0,07	0,47 *	0,49 *	0,36 *	0,37 *	0,34 *	

TABLE 13:   Intercorrelation Matrix for Asian Sugar Workers (n = 93)

* indicates correlations significantly different from 0
(p $\leqslant$ 0,05).

The complete intercorrelation matrix for 72 of the Black Sugar
Workers is displayed in Table 14.

	MATHS	SCIENCE	MA	HL FCT	A/3/1	BLOX	TRC	GFT
MATHS								
SCIENCE	0,85 *							
MA	0,40 *	0,45 *						
HL FCT	0,32 *	0,40 *	0,59 *					
A/3/1	0,26 *	0,39 *	0,53 *	0,31 *				
BLOX	0,04	0,11	0,24 *	0,49 *	0,25 *			
TRC	0,29 *	0,35 *	0,47 *	0,38 *	0,28 *	0,08		
GFT	0,26 *	0,40 *	0,19	0,15	0,27 *	0,38 *	0,25 *	

TABLE 14:   Intercorrelation Matrix for Black Sugar Workers (n = 72)

* indicates correlations significantly different from 0
(p ≤ 0,05)

As expected, the HL FCT correlated significantly with the MA
(0,61 and 0,59 for the Asian and Black groups respectively)
since the MA is a test of general mental ability. The tests
differ however in that the HL FCT is a non-verbal test and the
MA has a high verbal content.

The significant correlations between the HL FCT and the A/3/1
for both samples (r = 0,33 and 0,31 respectively) and between
the HL FCT and the BLOX for both samples (r = 0,55 and 0,49
respectively) reflect the spatial and reasoning components
common to the HL FCT, A/3/1 and the BLOX.

It is not surprising that there is a significant (although
moderate) correlation between the HL FCT and the TRC for the
two samples (r = 0,28 and 0,38 respectively) since the task
in the HL FCT involves reasoning and the TRC also has a reasoning
component.

Correlations between the HL FCT and the GFT were found to be significant only in the case of the Asian sample ($r = 0,49$) but not in the case of the Black sample ($r = 0,15$). Since the GFT can be thought of as a test of perceptual analysis, it appears that the Asian group may be using a similar analytic-type strategy in solving the problems of the GFT and HL FCT. Blacks, on the other hand, are probably using a more holistic, perceptual-type approach in solving the items of the GFT. Since this approach is not possible in the solution of most HL FCT items, the correlation between HL FCT and GFT is low.

4.3.2 SAMPLE vii

The intercorrelation matrix of the test battery scores of the entire group of Black male applicants ($n = 263$) is given in Table 15.

	IMA	HL FCT	A/3/1	BLOX	TRC	SL ART	F
IMA							
HL FCT	0,60 *						
A/3/1	0,41 *	0,35 *					
BLOX	0,29 *	0,36 *	0,20 *				
TRC	0,42 *	0,29 *	0,28 *	0,14 *			
SL ART	0,37 *	0,36 *	0,14 *	0,13 *	0,25 *		
F	0,27 *	0,25 *	0,12	0,23 *	0,06	0,22 *	
H	0,25 *	0,24 *	0,16 *	0,34 *	0,08	0,16 *	0,79 *

TABLE 15: Intercorrelation Matrix for Black Male Sample ($n = 263$)

* indicates correlations significantly different from 0 ($p \leqslant 0,05$).

	IMA	HL FCT	A/3/1	BLOX	TRC	SL ART	F	H	ELECTRONICS TEST	APPLIED MECHANICS TEST	TYPEWRITING TEST	ADDING MACHINE TEST
ELECTRONICS TEST	0,49*	0,54*	0,05	0,26*	0,18	0,11	0,01	0,08				
APPLIED MECHANICS TEST	0,32*	0,43*	0,26*	0,18	0,34*	0,11	-0,02	0,04	0,62*			
TYPEWRITING TEST	0,24*	0,29*	0,09	0,26*	0,09	0,08	-0,03	0,16	0,48*	0,42*		
ADDING MACHINE TEST	0,29*	0,30*	0,20	0,28*	0,26*	-0,13	-0,18	0,05	0,63*	0,48*	0,44*	
SELF STUDY	0,50*	0,51*	0,09	-0,31*	0,00	0,18	-0,17	0,06	0,42*	0,25*	0,30	0,62*

TABLE 16: Intercorrelation Matrix for 66 selected Black Male Technicians

From Table 15 it can be seen that the HL FCT correlates signifi-
cantly with the MA, A/3/1, BLOX and TRC. These correlations
are similar to those obtained for the Asian and Black Sugar
workers (see section 4.3.1).

The significant correlation between the HL FCT and the SL ART
(r = 0,36) is expected since both tests require a reasoning-type
approach in their solution.

Finally, the moderate but significant correlations between the
HL FCT and the two spatial tests, the F (r = 0,25) and the
H (r = 0,24) show that there is a spatial element shared by
these tests.

After selection, the sample size decreased to 66.

Intercorrelations between selected criterion scores and the test
battery scores are shown in Table 16.

From the previous matrix, it can be seen that the HL FCT correlates
substantially with all the criterion scores listed in Table 16
for the selected group of 66 trained male technicians. These
correlations are particularly remarkable considering the
highly selected nature of the sample.

In all cases, the HL FCT correlates more highly with the criterion
scores (ELECTRONICS, APPLIED MECHANICS, TYPEWRITING, ADDING MACHINE,
ELECTRONICS SELF STUDY) than does any other test in the battery (but
not always significantly more highly).

4.3.3 SAMPLE viii

The intercorrelation matrix obtained on the 199 Mabopani East
Technikon first year students is presented in Table 17.

	IMA	HL FCT	BLOX	TRC	SL ART	ESTIM
IMA						
HL FCT	0,48 ˣ					
BLOX	0,38 ˣ	0,35 ˣ				
TRC	0,26 ˣ	0,17 ˣ	0,04			
SL ART	0,46 ˣ	0,34 ˣ	0,15 ˣ	0,22 ˣ		
EST	0,28 ˣ	0,26 ˣ	0,07 ˣ	0,16 ˣ	0,24 ˣ	

TABLE 17: Intercorrelation matrix for 199 first-year Technikon students

As expected, the HL FCT correlates significantly with the IMA, BLOX, TRC as well as with the two new NIPR experimental tests, SL ART and ESTIM, which both require a reasoning-type approach in solving the items.

4.4 CONCLUSIONS

For the samples considered above, both the MA/IMA and the HL FCT correlate substantially with the criterion data. The HL FCT appears to perform better than the IMA as a predictor in the case of the selected group of Black male technicians. The two tests are by no means interchangeable since they are only moderately correlated (correlations vary between 0,48 for sample viii and 0,61 for sample v). It would appear that some of the predictive power of the MA might come from its verbal content whereas the ability of the HL FCT to predict criteria is probably based on the non-verbal conceptual demands which it places on the testee. It would therefore seem to be of value to employ both tests in many selection procedures.

REFERENCES

5.0   REFERENCES

CAROTHERS, J C (1953)   The African Mind in Health and Disease.
Geneva, World Health Organization.

CATTELL, R B (1971)   Abilities: Their Structure, Growth and Action.
Boston, Houghton Mifflin.

CIBOROWSKI, T and   Cultural differences in learning conceptual
M COLE (1971)   rules. International Journal of Psychology,
6, 25 - 37.

COULTER, M A (1977)   Item Response Evaluation. CSIR Special Report
PERS 186. Johannesburg, National Institute
for Personnel Research.

CRYNS, A G J (1962)   African Intelligence: a critical survey of
cross-cultural intelligence research in
Africa south of the Sahara. Journal of Social
Psychology, 57, 283 - 308.

DEREGOWSKI, J B and   Performance on a sorting task: a cross-cultural
R SERPELL (1971)   experiment. International Journal of Psychology,
6, 273 - 281.

HALL, B A (1978)   Short-term validation of a selection procedure
for student nurses for the general diploma in
nursing. CSIR Special Report PERS 318.
Johannesburg, National Institute for Personnel
Research.

HORST, P (1953a)   Correcting the Kuder-Richardson reliability
                          for dispersion of item difficulties.
                          Psychological Bulletin, 50, 371 - 374.

HORST, P (1953b)   Relationships between several Kuder-Richardson
                          reliability formulas. Educational and Psychological
                          Psychological Measurement, 13, 497 - 504.

IRWIN, H M and   Ability and preference in category sorting
D H McLAUGHLIN (1970)   by Mano schoolchildren and adults. Journal
                          of social Psychology, 82, 15 - 24.

JAHODA, G (1956)   Assessment of abstract behaviour in a non-
                          western culture. Journal of abnormal and
                          social Psychology, 49, 109 - 110.

KELLEGAHAN, T (1968)   Abstraction and categorization in African
                          children. International Journal of Psychology,
                          49, 109 - 110.

KUDER, G F and   The theory of the estimation of test reliability.
M W RICHARDSON (1937)   Psychometrika, 2, 151 - 160.

LÈVY-BRUHL, L (1932)   Primitive Mentality. New York, Macmillan.

LORD, F M and   Statistical theories of mental test scores.
M R NOVICK (1968)   Addison-Wesley, Massachusetts.

MAUGHAN-BROWN, E A   The NIPR Batch-Processing Programme System:
(1973)   Vol 3 - Item Analysis Programme (NP50).
                          CSIR Special Report PERS 211. Johannesburg,
                          National Institute for Personnel Research.

NATIONAL MANPOWER COMMISSION (1982)
Report of the National Manpower Commission for the period 1 January 1981 - 31 December 1981. Pretoria, Department of Manpower.

NUNALLY, J C (1978)
Psychometric Theory, New York, McGraw-Hill Book Company.

OKONJI, D M (1971)
A cross-cultural study of the effects of familiarity on classificatory behaviour. Journal of cross cultural Psychology, 2, 39 - 49.

POSNER, M I (1964)
Information reduction in the analysis of sequential tasks. Psychological Review, 17, 491 - 504.

PRICE-WILLIAMS, D R (1962)
Abstract and concrete modes of classification in primitive society. British Journal of educational Psychology, 32, 50 - 61.

SPENGLER, O (1926)
The decline of the West. London, Allen and Unwin.

TAYLOR, J M (1978)
Short-term validation of a selection battery for trainee engineering technicians. CSIR Special Report, PERS 270. Johannesburg, National Institute for Personnel Research.

TAYLOR, T R (1976)
Test administrator's Manual for Figure Classification Test. K7.57. Johannesburg, National Institute for Personnel Research.

TAYLOR, T R (1977)
The measurement of concept identification differences among literate Blacks. Psychologia Africana, 17, 31 - 47.

VAN RENSBURG, C (1975)   Stepping into the Future. Johannesburg,
                            Erudita - Publikasies.

VISSER, B L (1978)   Development of a testing programme for
                            Black university students for purposes of
                            selection and careers counselling. CSIR
                            Special Report PERS 281. Johannesburg,
                            National Institute for Personnel Research.

WERNER, H (1948)   Comparative Psychology of Mental Development.
                            Chicago, Follett.

APPENDIX

FREQUENCY DISTRIBUTION OF SCORES ON THE HL FCT
93 MATRICULATED ASIAN SUGAR WORKERS IN TECHNICAL FIELD
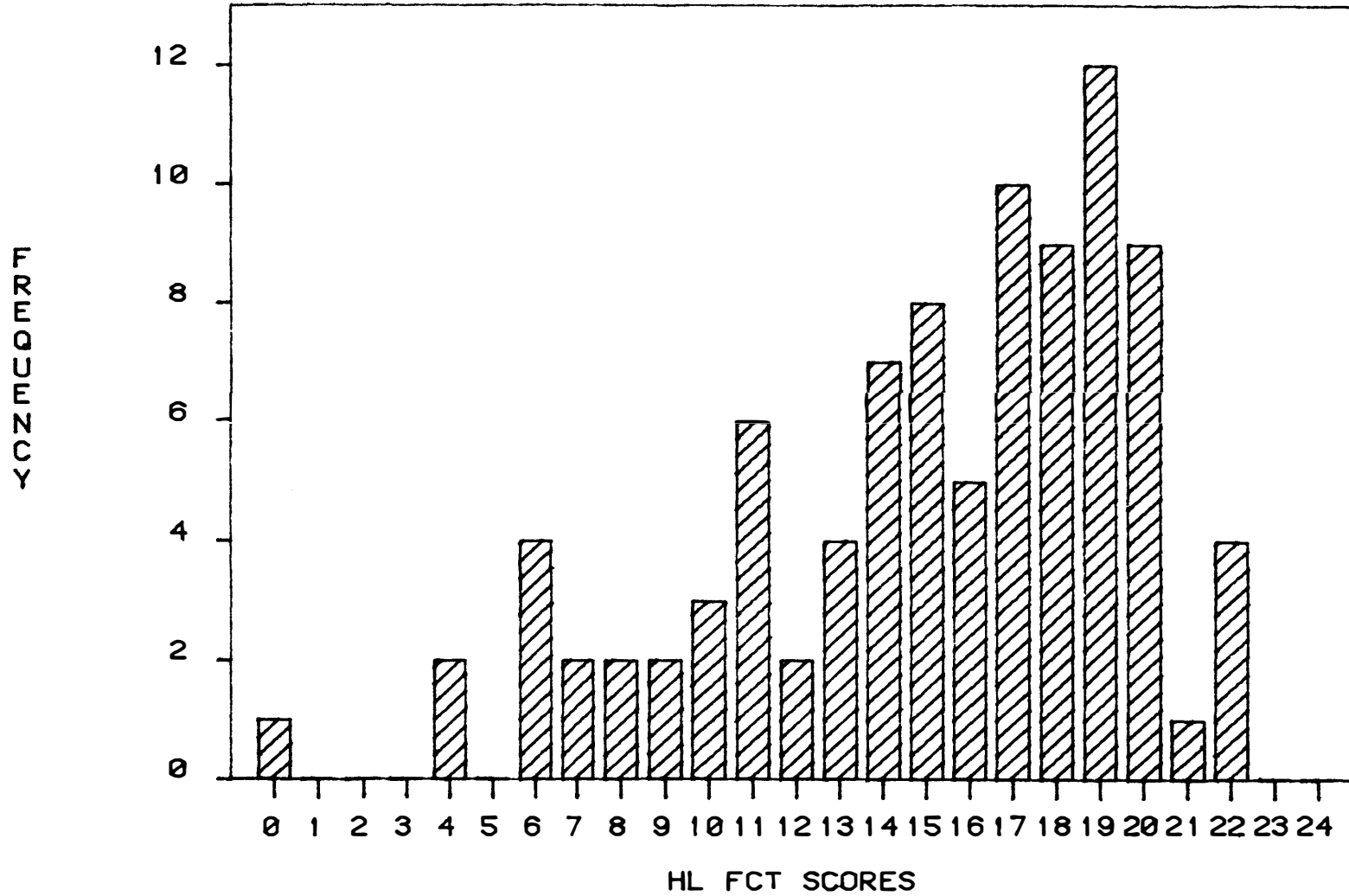
Figure 3

FREQUENCY DISTRIBUTION OF SCORES ON THE HL FCT
165 MATRICULATED BLACK SUGAR WORKERS IN TECHNICAL FIELD
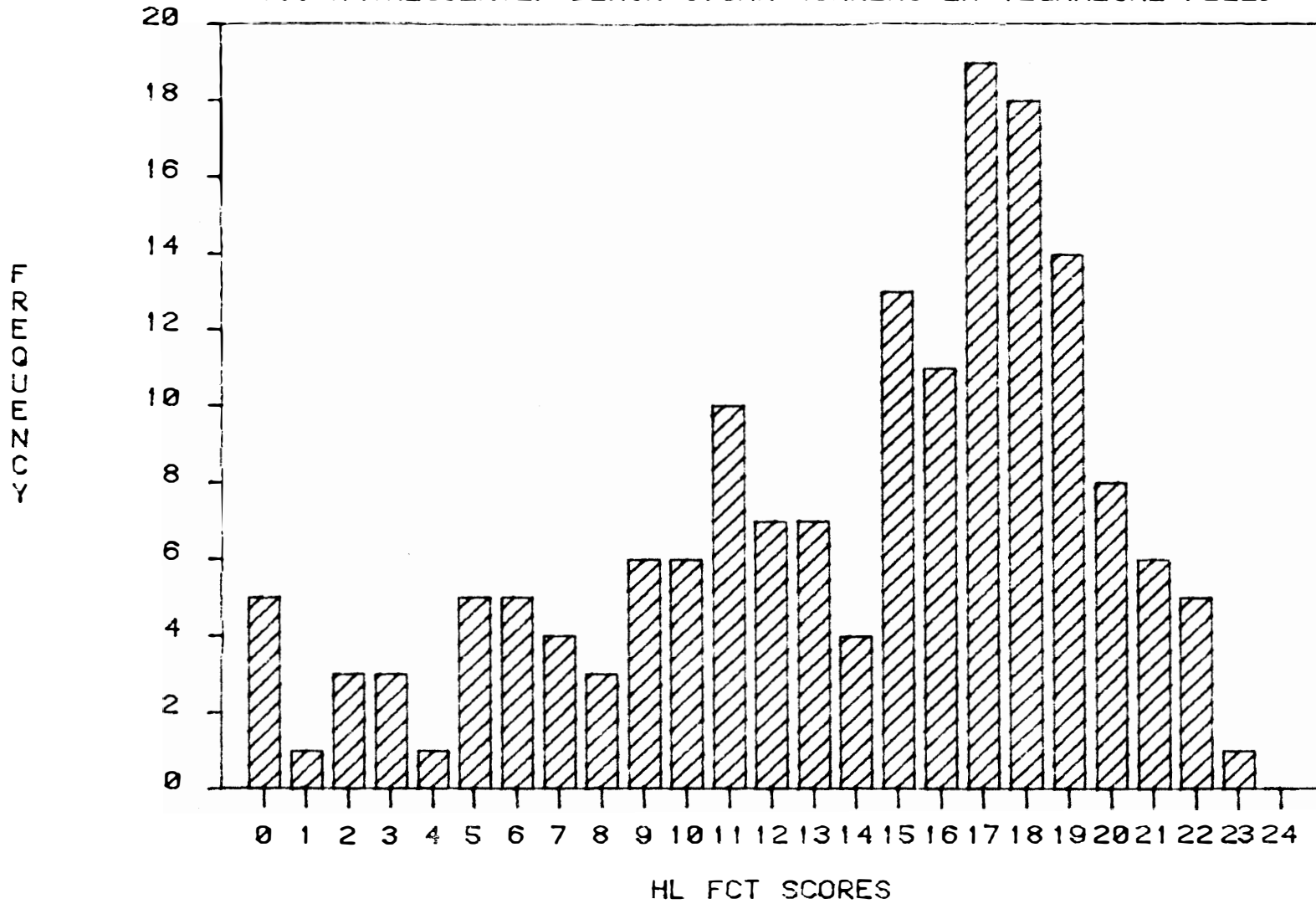
Figure 4

FREQUENCY DISTRIBUTION OF SCORES ON THE HL FCT
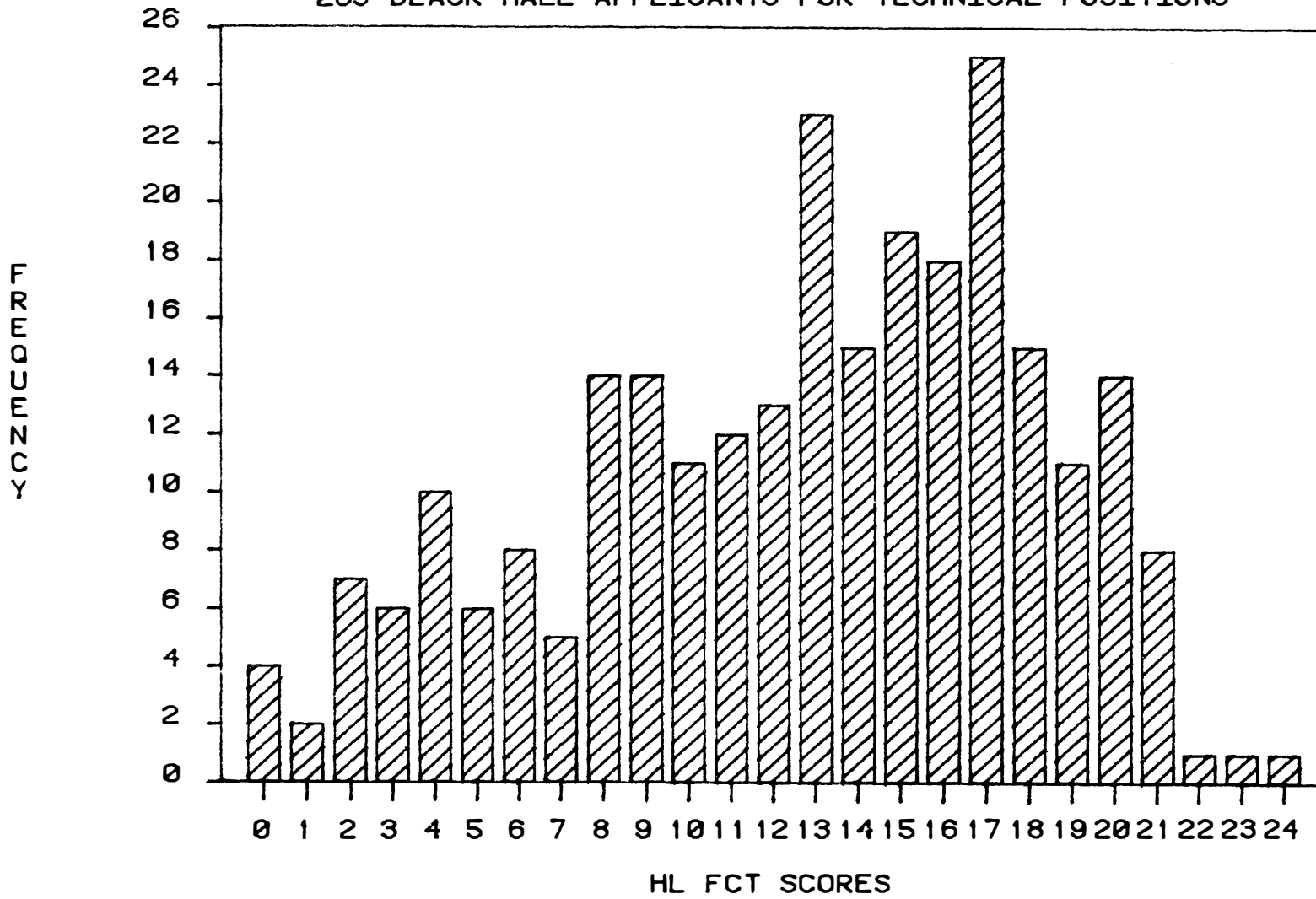263 BLACK MALE APPLICANTS FOR TECHNICAL POSITIONS

HL FCT SCORES

Figure 5

FREQUENCY DISTRIBUTION OF SCORES ON THE HL FCT
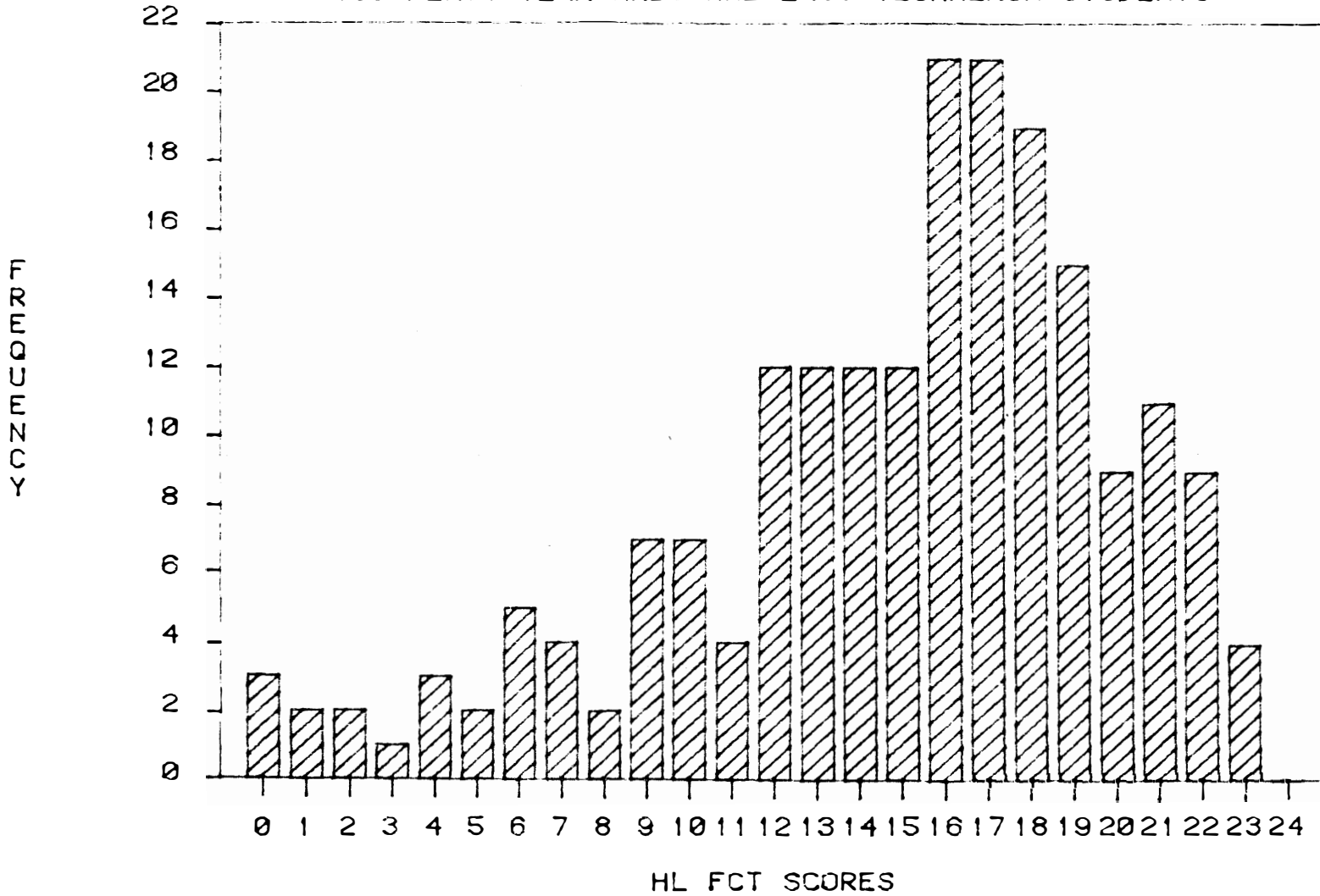199 FIRST YEAR MABOPANI EAST TECHNIKON STUDENTS

Figure 6

FREQUENCY DISTRIBUTION OF SCORES ON THE HL FCT
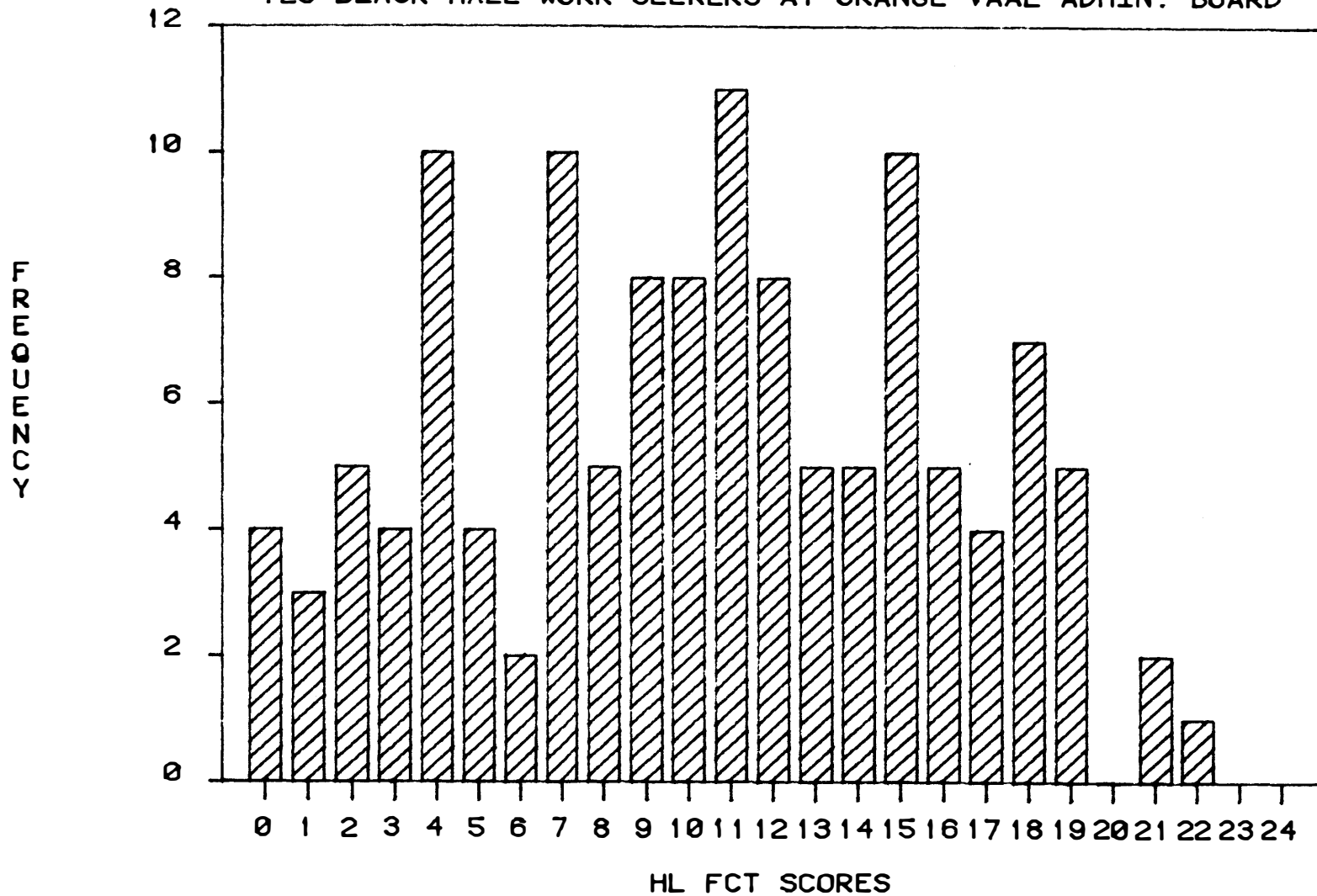126 BLACK MALE WORK-SEEKERS AT ORANGE-VAAL ADMIN. BOARD

HL FCT SCORES

Figure 7

127 MATRIC MALE SOWETAN PUPILS STUDYING MATHS AND SCIENCE

Figure 8