SCIENTIA
WNNR · CSIR

SPECIAL REPORT

PERS. 187

THE INTELLECTUAL STRUCTURE OF C.S.I.R. SCIENTISTS

C.S.I.R. SPECIAL REPORT NO. PERS. 187
U.D.C.5.007+6.007:159.928.22.073]:572.9.026](680)

Johannesburg, South Africa.
February, 1973.

THE INTELLECTUAL STRUCTURE OF C.S.I.R. SCIENTISTS

by

J.M. VERSTER

PSYCHOMETRIC DIVISION

NATIONAL INSTITUTE FOR PERSONNEL RESEARCH

COUNCIL FOR SCIENTIFIC AND INDUSTRIAL RESEARCH

JOHANNESBURG

FEBRUARY, 1973

ISBN 0 7988 0307 X

CSIR Special Report PERS 187

Published by
National Institute for Personnel Research,
South African Institute for Scientific and
Industrial Research,
P.O. Box 10319,
Johannesburg, Republic of South Africa.

February 1973

Printed in the Republic of South Africa by
National Institute for Personnel Research.

ACKNOWLEDGEMENTS

This project was directed by Mr D.J.M. Vorster, Director of the National Institute for Personnel Research. The research was carried out as part of the activities of the Psychometrics Division, under guidance of Mr D.W. Steyn.

It forms part of the following projects:

4384 Compilation of a high level differential test battery.

4386 Development of a high level logical reasoning test.

4374 Construction of a high level inductive reasoning test.

Permission to test a sample of C.S.I.R. research staff was granted by the President, Dr C. van der Merwe Brink.

The willing co-operation of Institute Directors, Divisional Heads and scientific research staff at the C.S.I.R. is gratefully acknowledged.

The research discussed in this report formed part of a thesis submitted to the University of South Africa in fulfilment of the requirements for the degree of Master of Arts in Psychology.

SUMMARY

This study was undertaken to fulfill two major aims. Firstly to determine the validity of the Thurstonian model of intellectual structure in the population of CSIR research scientists and secondly to determine the construct validity of the Deductive Reasoning Test.

A battery of 14 ability tests selected as references for factors of deduction, induction, spatial reasoning and verbal meaning was compiled and administered to a representative sample of 160 CSIR scientists. The results were analysed by means of factor analysis and multi-dimensional scaling procedures. Partial support for the Thurstonian model of differentiated Primary Mental Abilities was found. No deduction factor emerged. Deduction and induction tests loaded on a single dimension. Separate factors of verbal meaning and spatial reasoning were confirmed. The merging of deduction and induction tests on a single dimension was partially attributed to the low reliabilities of the reference tests and partially to the differentiation-integration theory of mental abilities proposed by Lienert and Crott (1964).

It was recommended that the Thurstonian model of intellect be used as a theoretical guide-line in the development of a differential ability selection battery for use in the CSIR.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Recent trends in psychological measurement at the NIPR point to a growing faith in the theory of a differentiated profile of human behaviour. The theory contends that, as the organism evolves, a number of conceptually and behaviourally distinct constructs emerge from the broad, underlying domains of personality, intelligence and motivation. The theory is backed by an impressive volume of psychometric research and lends itself admirably to psychometric methodology. It also has a natural appeal to the personnel psychologist since it agrees well with his conceptualization of multidimensional criteria. Conclusive evidence of the validity of the theory is still forthcoming and it should be treated with caution in the interim. An overzealous interpretation of the theory could lead to an unwarranted proliferation of trivial measures and consequently to an artificially inflated body of research in support of an unrealistically atomistic conceptualization of human behaviour. If properly interpreted, differentiation theory provides a useful and valid way of thinking about behavioural constructs provided that the ultimately global, integrated nature of human functioning is never lost from sight.

In the present study an attempt is made to find empirical justification for conceptualizing the human intellect in terms of a differentiated profile of constituent abilities. This view not only reflects current NIPR thought, but is in keeping with the well-known work of Thurstone[1),2)] (1938, 1940) on the Primary Mental Abilities of man. The basic theoretical rationale underlying this view is that intelligent behaviour can be meaningfully classified into a limited number of broad, operationally defined constructs. Each of these is assumed to be amenable to independent measurement by means of an appropriate ability test. Each measurable ability is regarded as a continuous psychological dimension underlying reliable individual differences. The development of abilities is ascribed to an ongoing interplay between genetically determined attributes and environmental influences during the growth of the organism. Hereditary factors as well as continuous pressures from the environment are therefore responsible for the patterning of abilities in an individual and for the structure of intellect in a culture.

2. DESIGN OF THE STUDY

The study is undertaken in the population of graduate research scientists at the CSIR. The individuals comprising this population may be thought of as representing the upper end of the intelligence curve of the general population in the country. These individuals are in the vanguard of the nation's scientific, industrial and technolcgical endeavour. Among their ranks are the scientific pioneers of to-day and the leaders of the industrial community of tomorrow. They constitute an indispensible factor in the growth of the country's economy. The need to develop accurate and reliable instruments for the identification of individuals who potentially belong to this population cannot be too greatly stressed. Yet to date little systematic research has sought to clarify the nature of intellectual functioning and the composition of intellectual structure of this highly sophisticated sample. The present study constitutes an attempt to meet this need.

The major aim underlying the study is to determine whether the Thurstonian theory of Primary Mental Abilities provides an appropriate model for describing the intellectual structure of the population of scientists at the CSIR. The study is, in this sense, a test of the generality, and hence the validity of the Thurstonian model. If the validity of the model can be demonstrated, the

Thurstonian theory will be adopted as a frame of reference for the construction of a new differential ability test battery to be used in the selection of applicants seeking scientific research posts at the CSIR.

The study is conducted along the lines of traditional Thurstonian factor analytical methodology (Thurstone,[3] 1947). A test battery comprising reference tests for four postulated Thurstonian Primary Mental Abilities considered of relevance to scientific research work is compiled and administered to a representative sample of CSIR scientists. The results are treated by means of factor analysis. An inter-group factor analysis procedure is used to compare the factor structures obtained for separate English- and Afrikaans-speaking samples.

The experimental aims and design of the study include a simultaneous evaluation of the preliminary Deductive Reasoning Test (D.R.T.) developed at the NIPR. Three preliminary versions of the test, designed according to an explicitly formulated conceptual model of deductive processes, were prepared (Verster,[4] 1970). A comprehensive analysis was made (Verster,[5] 1972) of the respective contributions of the parameters of the test model to item difficulty and the results of the investigation were utilized in the construction of a preliminary 45 item version of the test.

By including this test in the test battery in the present study, it will be possible to determine the construct validity of the 45 item D.R.T. This is the second major aim of the study and the tests in the reference battery are selected with this consideration in mind.

2.1. Selection of Reference Abilities

The question of which abilities to measure in the present study was necessarily guided by practical considerations. While it is of considerable scientific interest to test the invariance of the

Thurstonian model in a high level population such as that represented by CSIR research scientists, it should be borne in mind that the major practical consideration underlying the investigation is to uncover the nature of those mental abilities that are likely to be of greatest relevance in the selection and assessment of research scientists.

It was decided at the outset, therefore, that it would be uneconomical to attempt to include reference tests for all of Thurstone's identified Primary Mental Abilities (P.M.A.'s). The study indeed, is essentially a test of the theoretical nature of Thurstone's model; of his hypothesis that the abilities of man are few and fundamental, susceptible to operational definition in behaviourally relevant terms and that they may, at times, be partially correlated with one another. The aim of the investigation is not to demonstrate the existence of each of the Primary Mental Abilities that Thurstone postulates. Moreover, it should be remembered that deductive reasoning ability, which is of central interest in the present study, was not actually included among the final factor abilities which Thurstone described as the Primary Mental Abilities of man (Thurstone and Thurstone,[6)] 1941). Apart from deduction however, the remaining reference abilities in this study are selected from among those described by Thurstone. Their inclusion in the reference battery was determined by considerations of their relevance to scientific research activities. The absence of any good validation studies in this respect necessitated a subjective criterion evaluation. It is considered unlikely though, that the relevance of the four selected abilities to the intellectual demands of a scientific research environment will be disputed.

Apart from the interest in measuring an ability of deduction, considerable effort was spent in finding suitable reference tests for a factor of induction, for this ability is also of crucial interest to the design of the project. It is of particular interest to see whether separate factors for induction and deduction can be produced. The reasoning processes represented by these two intellectual abilities are easy to distinguish conceptually. But whether they remain operationally distinct and represent separate psychological abilities is not so clear. The position adopted in the present study is that the mental operations called for in deduction are slightly different from those required in induction and that the two processes may well underlie separate abilities. It is expected however, that these abilities will be highly related and may even operate in conjunction with one another so that if separate factors can be produced at all they will be highly correlated. In addition the issue is likely to be confused further by a lack of suitable tests.

In his original investigation Thurstone[7)] (1938) found encouraging evidence for the existence of an independent induction ability. His tests for this factor all required the use of a rational thinking strategy to determine the rule which, from an implicitly provided set of culling rules or principles, best succeeded in accounting for the patterned organisation of data within each item. This type of rule discovering or rule identifying behaviour is easy to distinguish conceptually from that required in the deductive tests that Thurstone employed.

Guilford[8)] (1967) does not even presume the existence of an induction ability as a separate factor. He prefers to regard inductive variance as being rightfully shared out among several of his more specific factor abilities. French et al.[9)] (1963) view induction as a complex second order factor ability underlying several, as yet unidentified first order abilities. This view, indeed, seems the one best supported in the literature on induction.

It is possible that the problematical nature of inductive reasoning can be resolved by making a conceptual distinction between different classes of induction. A useful distinction that has been described more fully elsewhere by the author (Steyn and Verster,[10] 1972), is that suggested by Kneebone[11] (1963, p. 366) between conceptual induction and determinative induction. The former relates to the type of thinking involved in the production of scientific theories. The products of conceptual induction not only add to the mass of available knowledge, but also make possible a more adequate understanding of already recorded facts by changing scientific perspective. Determinative induction, on the other hand, makes an addition to the content of available knowledge but leaves the form, or the scientific perspective, unchanged. Determinative induction is dependent on a statistical principle of optimum choice. From a given range of specified possibilities the one that is most appropriate in the light of available evidence is selected. This is the kind of inductive reasoning that is more commonly encountered in scientific research and is considered worthy of measurement. Conceptual induction, on the other hand, is not believed to be amenable to measurement with the traditional type of cognitive test. In a related NIPR project on which the author is engaged, a test of determinative induction will be developed. To date, however, no pure tests of this kind of reasoning are available and the best of the available tests of complex induction will therefore, out of necessity, be used in the present study.

Attempts will also be made to measure a dimension of spatial reasoning ability. Spatial reasoning or simply space was one of the first abilities to be identified unambiguously in the P.M.A. investigations. This factor, according to Thurstone[12] (1941), is found in tests that require the subject to manipulate objects imaginally in two or three dimensions. It is quite distinct from perceptual processes which only require the perception of detail in a flat surface and which do not require the imaginal movement of an object in two or three dimensions. Thurstone[13] (1941) also raised the important psychological question of whether the space factor is somehow related to kinesthetic imagery. This point however, has not yet been clarified. Neither has it been shown that space perception enters into auditory material. As it stands at present, the factor is characterised as facility in spatial and visual imagery.

French et al.[14] (1963) also recognised a space factor called Spatial Orientation, which is defined as "the ability to perceive spatial problems or to maintain orientation with respect to objects in space". Tests loading on spatial orientation seem to involve perception of the position and configuration of objects in space. Whether the material involved is presented in two dimensions or three seems to have little effect on the nature of this factor. It is thus essentially the same factor as Thurstone's space factor and indeed, some of Thurstone's tests are used by French et al.[15] (1963) to identify the factor.

Following the distinction made by French et al.[16] (1963) between the psychological nature of spatial orientation and visualization, the space factor in the present study will be denoted spatial reasoning. Tests loading on spatial orientation require a degree of manipulative reasoning in response to the visual objects as a whole, whereas in visualization tests the emphasis is on the perception of visual detail without resorting to an imaginal manipulation of the material.

In addition to factors of induction, deduction and spatial reasoning, a factor of verbal meaning will be included in the present study. Comprehension of verbal material is central to all fields of scientific research endeavour. Thurstone[17] (1941) found a distinction between two major kinds of verbal factors. One was a verbal meaning factor, the other a factor of verbal fluency. It is the former of these that is important here. Although this sort of factor is probably general to all language groups, its stability depends on the familiarity of the population tested with the language medium in the tests. In order to measure this factor in the CSIR population it will consequently be necessary to use separate tests for English- and Afrikaans-speaking subjects.

The four selected primaries, deduction, induction, spatial reasoning and verbal meaning are considered the most central and relevant abilities to scientific research work. Certain other Thurstonian primaries such as verbal fluency, perceptual speed, memory etc., are likely to be less generally employed in all branches of science. The mathematician, for example, makes little demand on verbal fluency skills, whereas certain of the applied sciences, for example, mechanical engineering, do not call for special abilities in speed of perception. In addition, there is much controversy over the psychological nature of speed tests while many psychologists would argue that fluency measures belong more properly in the domain of creativity, or divergent production. It is possibly true that a good memory is a useful asset to any research scientist, but a memory factor has been excluded from the present investigation for two reasons. Firstly, most factor analytical studies of the cognitive domain have shown the memory factor to be relatively uncorrelated with other abilities. This implies that efficient intellectual functioning in other areas is not necessarily dependent on a good memory. If an individual memory factor were to emerge in yet another study little in the way of a new understanding of the structure of intellect would follow from this finding. Secondly, it is argued that there are few good tests of memory presently available. Most of the currently employed tests for this factor rely on rote memory or memory for meaningless material. It would be more valuable to have a measure of memory for meaningful stimulus material.

The memory domain thus appears to warrant further independent investigation along fresh avenues.

It is expected that the four abilities selected for this study will be mutually intercorrelated since a common component of reasoning is involved in each. Indeed, it is likely that a second-order analysis would produce a single general factor identifiable as some form of general reasoning ability. Nevertheless, it will be of considerable interest to see whether separate first-order primary factors can be produced in a battery of tests all of which require some form of reasoning activity.

2.2. Compilation of a Reference Battery

Following accepted factor-analytical theory the battery contains sufficient reference tests to allow at least three markers for each expected factor. Four postulated tests for deduction and four for induction are included in order to determine these factors more clearly, if they emerge at all. Three tests for each of the remaining factors are considered sufficient since the psychological nature of these abilities is more clearly understood.

2.2.1. Reference Tests for Deduction

The anticipated difficulty in finding suitable reference tests for deductive reasoning proved justified. The ambiguity surrounding the nature of this factor has already been discussed elsewhere by the author (Verster[18], 1970). It was pointed out that tests of syllogistic reasoning have consistently proved the most stable measures of this ability to date. Furthermore, as mentioned previously (Verster[19], 1970), non-verbal tests of deduction have been most difficult to locate. The writer was unsuccessful in obtaining a copy of Blakey's[20] (1941) non-verbal test of Form Reasoning which would have been useful. Canisia[21] (1962) also claims to have developed a non-verbal test of deduction, but this too could not be obtained. Other non-verbal tests involving deduction (Botzum,[22] 1951) do not seem to be factorially stable since they have at times shown substantial loadings on other factors (French et al.[23] 1963).

The Figure Series Test of the Senior Aptitude Test Battery (Fouche and Alberts[24] 1970) of the Human Sciences Research Council (H.S.R.C.) was thus selected as a non-verbal reference test for deduction. The format of this test is similar to that of Blakey's[25] (1941) Form Reasoning.

Verbal references for this factor were somewhat less difficult to find. Special care was taken not to include other tests of formal syllogisms, since this would spuriously inflate correlations with the D.R.T. but would contribute little to a fuller identification of the factor.

The four postulated tests for deduction are the following:

Deductive Reasoning Test	(D.R.T.)
Figure Series	H.S.R.C. Senior Aptitude Tests
Reasoning Ability	NIPR General Selection Battery
Inference	E.T.S. French et al. (1963)

In Reasoning Ability the item format is in the form of a problem followed by two statements of alleged fact having apparent relevance to the problem. The subject is required to indicate which of five possible combinations of the two fact statements, that can be considered either independently or in conjunction with one another, will provide a solution to the problem. Actual answers to the problems are not required.

The item format in Inference requires the subject to indicate which one of five possible conclusions follows logically from a given statement.

2.2.2. Reference Tests for Induction

French et al.[26] (1963) define induction as associated abilities involved in the finding of general concepts that will fit sets of data, the forming and trying out of hypotheses.

Although their definition is somewhat loose and vague, it is clear that these writers view induction as a multi-dimensional construct. In fact, they prefer to regard induction as a second-order factor underlying several more distinct primary factors that have not yet been clearly defined. They argue, however, that the three reference tests suggested in their Kit (French et al.[27] 1963) have been found to correlate with each other and will define a factor representing induction when it is desired to separate inductive variance from factors in other areas. Since this coincides with the purpose in the present study their guide will be followed in the absence of better information about induction.

The three tests which they propose represent three content areas, letters, patterns and numbers and, in Guilford's terminology, two kinds of products, classes and systems.

To give more stability to the induction dimension, a fourth test was added to the three suggested by French et al.[28] (1963). This is Pattern Completion, a test from the Senior Aptitude Test Battery of the H.S.R.C. (Fouche and Alberts,[29] 1970). It is a matrix test in which rules have to be inferred or induced from the given elements of the matrix.

The four reference tests for inductive reasoning are:

Letter Sets	E.T.S. French et al. (1963)
Locations	E.T.S. French et al. (1963)
Figure Classification	E.T.S. French et al. (1963)
Pattern Completion	H.S.R.C. Senior Aptitude Tests

Letter sets was suggested by Thurstone's Letter Grouping. The item format consists of five sets of four letters each. The task is to find the rule that relates four of the sets to each other and to mark the one which does not fit the rule.

Locations was suggested by Thurstone's Marks. In each item five rows of places and gaps are given. In each of the first four rows, one place in each is marked according to a rule. The subject is to discover the rule and to mark one of the five numbered places in the fifth row accordingly.

Figure Classification is adapted from Thurstone's test of the same name. Each item presents two or three groups each containing three geometrical figures that are alike in accordance with some rule. The second row of each item contains eight test figures. The task is to discover the rule and assign each test figure to one of the groups. There are twenty-eight items with eight test figures each, resulting in 224 possible responses to score.

2.2.3. Reference Tests for Spatial Reasoning

In selecting reference tests for this factor the guide offered by French et al.[30] (1963) was again followed, but an NIPR test, Blox, was substituted for the Spatial Orientation test of Guilford which French et al. recommend.

The three tests for this factor are:

Card Rotation	E.T.S. French et al. (1963)
Cube Comparison	E.T.S. French et al. (1963)
Blox	NIPR Perceptual Battery

Card Rotation was suggested by Thurstone's well-known Space test Cards. In each item a drawing is given of a card cut into an irregular shape. To its right are six other drawings of the same card sometimes merely rotated to different orientations and sometimes turned over onto the other side. The subject is to indicate which ones show the card not turned over.

Cube Comparison was adapted from Thurstone's Cubes. In each item a drawing of a cube is presented. Assuming no cube can have two faces alike, the subject is to indicate which items present drawings that can be of the same cube and which present drawings that cannot be of the same cube.

Blox is a sub-test of the NIPR Perceptual Battery. On each page five sets of blocks or three dimensional cubes are presented. The number of blocks within each set varies between two and six and these are arranged together to form various random patterns. The subject has to analyse these patterns or geometric configurations and identify them when presented in different orientations in the items below. There are six items for each set of five stimulus configurations on a page and a total of forty items in the test.

2.2.4. Reference Tests for Verbal Meaning

A verbal meaning factor has been found in at least 70 published studies (French et al.[31] 1963). It seems likely that several linearly independent sub-factors can be found in the verbal domain if the specialized vocabularies of certain professions, occupations or dialectical groups are tested. Reference tests of the verbal factor consequently should be tests which sample features of a particular language very widely. According to French et al.[32] (1963) vocabulary tests are usually more desirable than tests of grammar and other language features because they are easier to construct and administer. On the whole their loadings on verbal meaning are higher and they are less likely to have loadings on other factors. French et al.[33] (1963) suggest five different vocabulary tests as reference markers for this factor.

Following Thurstone[34] (1938) it was considered desirable to include a test of reading comprehension as a marker for this factor although reading comprehension has also been shown to contain general reasoning.

A special problem regarding the measurement of verbal meaning in the present study relates to the bilingual research population at the CSIR. Clearly this ability should be measured in the home language of the subject concerned. It was realized at the outset, therefore, that separate tests for English- and Afrikaans-speaking subjects would have to be used. For this reason, tests involving grammar, analogies, synonyms and antonyms, etc., were avoided as their inclusion could lead to insurmountable problems in attempting to achieve comparability of difficulty, complexity and construct validity across the two languages.

In the light of this problem, and in accordance with the recommendations of French et al.[35] (1963) it was decided to use more than one vocabulary test, since these are easiest to construct in the different languages, together with a good test of reading comprehension which could be translated into both languages, as references for verbal meaning.

The NIPR High Level Battery provided a suitable test of Reading Comprehension which is available in both languages and has been used for many years in the selection of high level personnel. Tests of English Vocabulary and Afrikaans Woordeskat were also taken from this battery. A second test of English Advanced Vocabulary was taken from among the tests recommended by French et al.[36] (1963) but a suitable Afrikaans counterpart could not be obtained. A new test of Gevorderde Woordeskat was therefore specially prepared by the author for use in this project.

Very briefly, the test was constructed in the following way to achieve relative comparability of difficulty and content with the

E.T.S. Advanced Vocabulary test. The stimulus words from the latter test were classified into frequency categories according to a recent English word count (Wright,[37] 1965). It was found that about 80% of these fell into the least common frequency category. A comparative Afrikaans word count (Die Afrikaanse Woord-telling[38] - Opgestel deur die Nasionale Buro vir Opvoedkundige en Maatskaplike Navorsing, 1958) was obtained and 50 words were selected from the corresponding least common frequency category. In selecting words, mainly adjectives and adverbs were chosen for each stimulus word. The preliminary Gevorderde Woordeskattoets was used with the understanding that an item analysis would be undertaken to purify the test of poor items before using it in any correlational analyses. Hence on the expectation that about 20 items might be dropped after the item analysis, the test was made sufficiently long to ensure that at least 30 good items remained in the final test. It will be seen from the analyses in the following chapter that the final "purified" 30 item version of the test had a reliability coefficient in the vicinity of 0,9 so that it seems this strategy proved successful.

The final test battery, the test time limits, number of items per test and predicted factor loadings are given in Table 1. It should be pointed out that the prescribed procedure for administering the E.T.S. reference tests was not followed. These tests are constructed in the form of two short parallel halves each with a separate time limit. But in view of the fact that the total test battery was already time-consuming, it was decided to economise on time by ignoring the break between the two halves in these tests. Single time limits were imposed and these are reflected in Table 1.

TABLE 1

TEST BATTERY

Test	Time	No. of Items	Factor	Source
LOCATIONS	14	28	I	ETS
D.R.T.	40	45	D	NIPR
CARD ROTATION	12	28	S	ETS
H.L. VOCABULARY	12	40	V	NIPR
LETTER SETS	16	30	I	ETS
INFERENCE	14	20	D	ETS
CUBE COMPARISON	10	42	S	ETS
H.L. READING COMP.	20	20	V	NIPR
FIGURE CLASS.	18	224	I	ETS
BLOX	30	45	S	NIPR
REASONING ABILITY	30	30	D	NIPR
PATTERN COMP.	10	30	I	HSRC
FIGURE SERIES	10	30	D	HSRC
ADV. VOCABULARY/ GEV. WOORDESKAT	10/ 16	30/ 50	V	ETS/ NIPR

I = Induction

D = Deduction

S = Spatial Reasoning

V = Verbal Meaning

2.3. The Sample

The aim in drawing the sample was to obtain full representation of the population of research personnel at the CSIR. The findings that emerge from the analysis of the factor battery will determine which abilities are included in a differential test battery to be developed for use in the selection of CSIR research scientists. Since this assessment battery is to be used widely in the selection of applicants for every CSIR Institute and for all levels of status rank, it is desirable that the experimental sample from which the results are obtained should be as representative as possible of the total CSIR research population. Rank and institute were identified as the most relevant determining population variables and the sampling method employed is consequently designed to give proportionate representation across these two criterion variables.

Due to practical considerations, only 13 CSIR institutes situated in Pretoria and Johannesburg were included in the target population. Included among these are research institutes for physics, chemistry, mathematics, electrical engineering, mechanical engineering, building research, road research, water research, telecommunications research and the institute for Information and Research Services (IRS). The personnel research institute (NIPR) in which the author is employed was excluded from the sampling universe as many of the tests in the battery are known to the staff members of this institute.

All technical and other non-research ranks were excluded from the target population with the exception of certain graduated personnel from IRS, who are engaged full-time on research, but who for administrative reasons, are graded in the technical ranks. Only three individuals with technical ranks were included in the final sample. All other members of the sampling population come from the ranks of Assistant Research Officer (ARO), Research Officer

(RO), Senior Research Officer (SRO) and Chief Research Officer (CRO). The most senior research rank grading, Senior Chief Research Officer (SCRO), was excluded from the sample as many of the senior executives holding this rank would not have been available for testing had they been included. In addition, most SCRO's are engaged in managerial capacities rather than being actively involved in research.

The following characteristics define the population from which the sample was drawn.

Age Range	:	20 - 50 years
Sex	:	male and female
Education	:	Bachelors Degree (minimum)
Rank	:	ARO, RO, SRO, CRO
Institute	:	IRS, NPRL, NRIMS, TRU, NEERI, NMERI, NBRI, NIWR, NITR, NLRL, NIDR, NIRR, NFRI.

A pseudo-random sampling technique designed to give proportionate representation across ranks and across institutes was employed. Lists were obtained from each of the 13 participating institutes of the names of staff members who complied with the above specifications. The total number of research staff in the population was 345. The sample was drawn by ordering the names alphabetically within each institute's list according to rank and excluding every third name that occurred. This gave a two-thirds sample of 230 research workers with proportionate representation of rank and institute. In preference to merely specifying the requisite number of incumbents from each cell in the two-way rank by institute sampling matrix, subjects included in the sample were specified by name. This ensured greater randomness and hence better representation. If only the requisite number from each cell were stated and the actual allocation of subjects to the sample was left to the discretion of an institute director or divisional head, unwanted biases could enter the sample.

All individuals selected for the sample were provided with a circular informing them of the broad aims of the project and their co-operation was requested. Anonymity with regard to test results was guaranteed and the voluntariness of participation was stressed. Only 30 of the sampled subjects could or would not participate. The final sample size was thus exactly 200.

2.4. Administration of the Test Battery

2.4.1. Testing

For numerous practical reasons testing was arranged to occur in several separate test sessions over an extended period of about two months. A provisional testing time-table was prepared in collaboration with the Directors of the 13 CSIR institutes concerned and this was circulated to the subjects, who indicated which date would be most convenient for them to be tested. Alternative venues were provided for each date. It was stressed that an attempt would be made to test as many subjects as possible from the same institute on a particular date in order to cause least inconvenience to the institute concerned.

Actual group sizes tested on any particular date varied considerably but were kept below a maximum of 30 and were usually around 10 - 12 for ease of administration. Test venues differed from day to day depending on which venue was most convenient to the subjects being tested. All testing venues were modern and well-equipped and complied with the standard requirements for adequate test administration.

The total test battery took about five and a half to six hours to administer. Reasonable periods of rest between tests were allowed during this time and two short breaks, one for tea and one for lunch were provided.

The tests were administered strictly in the order they appear in Table 1 and the time limits were rigidly adhered to.

It will be seen that the D.R.T., which is the crucial test in the battery, was administered second. This meant that subjects were still fresh when doing this test while having had a chance to accustom themselves to the atmosphere in the testing room on the first test.

The Advanced Vocabulary test and the specially constructed Gevorderde Woordeskattoets were administered last as these tests had different time limits.

2.4.2. General

It was stressed from the outset that anonymity regarding the test results would be preserved. Subjects were told that all tests used were experimental and that no normative data for the evaluation of individual test scores was available. Raw scores on tests would consequently be meaningless and would under no circumstances be made available to subjects. It was furthermore explained that the test results would not be dealt with individually in the analyses, but would be treated correlationally in an attempt to determine the nature and order of the correlations among tests. In an attempt to arouse good test taking attitudes and strong motivation, it was stated that the mean differences between institutes and between ranks would be examined. Each subject was made to realise that he represented a limited group of subjects comprising his particular cell in the rank by institute matrix and that if he failed to take the tests seriously his poor performance would be reflected in the lower test means of the cell to which he belonged. This strategy proved successful in that it enabled subjects to identify with the project more strongly, and their motivation, in general, was extremely high throughout the lengthy test battery.

It was pointed out that, in view of the way the results were to be treated, some consistent form of identification would be required on each test cover. Since anonymity was to be preserved subjects were asked to identify their tests by indicating staff numbers on the covers. Subjects who refused to comply with this request were permitted to use any random pseudonym they wished. On the first test in the battery subjects were also asked to indicate their rank, institute, highest educational degree, date of birth (age), and branch of science in which they were currently working. This information was used as a check on the sampling procedure in order to detect whether any sampling biases were present.

3. RESULTS

3.1. Preparation of the Data

The tests were scored manually and the raw scores of the 200 subjects were recorded on a special data chart. The chart was carefully scrutinized in an attempt to detect anomalies such as missing information, or unusable test scores. It was discovered that five of the subjects had failed to understand the instructions in certain of the tests and their scores on these tests were consequently not valid. These subjects were rejected from the sample since their results were not usable.

A more serious anomaly that was apparent even at a cursory visual inspection of the test scores, was that a small number of subjects consistently scored well below the sample means on all tests. The discrepancies between the scores of these subjects and the mean scores of the total sample were particularly evident in the case of the verbal tests and especially on the two tests of vocabulary. On investigation it was found that these subjects were all foreigners, in the sense that their home language was given as neither English nor Afrikaans. It was found that subjects classified as foreign in terms of this criterion formed about 17% of the total sample. This figure is an accurate reflection of the proportion of foreign research staff at the CSIR. It was considered advisable to omit all foreigners from the sample before proceeding with the analyses.

The total usable sample was thus reduced to 160. This number was regarded as adequate for the purpose of the intended analyses.

The data from the reduced sample was accordingly loaded into the computer and machine scored as a further precaution against errors.

3.2. Test Characteristics

Before commencing with the analyses of the test battery it was necessary to evaluate the item characteristics of certain individual tests. Firstly, a detailed account of the 45 item version of the preliminary D.R.T. will be given, followed by a description of the final 36 item version of this test.

Consideration will also be given to the metric properties of the Gevorderde Woordeskattoets (G.W.S.) that was specially constructed by the author for this project.

An analysis of the items in the Reasoning Ability Test (R.A.T.) was performed in an attempt to discover the cause of the test's restricted variance and poor reliability. Poor items were eliminated from the test in an effort to increase the reliability of the instrument before incorporating it in the analysis of the test battery.

Locations, which was administered as the first test in the battery, proved to be too highly speeded for the CSIR sample. Some attempt was therefore made to partial out the undesired effects of speed on this test.

3.2.1. Test Characteristics of the Deductive Reasoning Test

3.2.1.1. Test Statistics

On the basis of the scores obtained by the total sample of 160 CSIR research scientists, the 45 item version of the D.R.T. was found to have a mean of 31,75 and a standard deviation of 8,12. Although the variance, and hence the discrimination power of the test is good, it appears that the scores were slightly negatively skewed in the sample tested. This could well be due to the fact that the most difficult items in test X, Y and Z were not included in this version of the test. It is possible that these items would have had greater variances in the CSIR population. Nevertheless, the spread of test scores was most satisfactory as evidenced by the high observed reliability coefficients.

Four different estimates of reliability were calculated for the 45 item version of the D.R.T. These are shown in Table 2. It will be seen that there is a high agreement among them. The respective assumptions underlying each of the modifications of the original Kuder-Richardson formula 20, and their implications in terms of mental test theory, are outlined in an article by Coulter[39)] (In Press).

TABLE 2

RELIABILITY ESTIMATES OF THE D.R.T.

KR_{21} *	0,93
KR_{20} (1)	0,91
KR_{20} (2)	0,89
KR_{20} (3)	0,90

KR_{21} *	:	KR_{21} with Tucker's correction for uniform (0,1) distribution.
KR_{20} (1)	:	no response taken as omission.
KR_{20} (2)	:	corrected for omission assuming uniform distribution of omissions.
KR_{20} (3)	:	no response taken as wrong.

These very high internal consistency coefficients justify the prediction of the Spearman-Brown extension formula that was applied to the three experimental versions of the test (Verster,[40] 1972). It is indeed encouraging to note that the test is capable of discriminating with such a high degree of consistency among research scientists. It will be seen later that apart from the Gevorderde Woordeskattoets, none of the other tests in the battery has the same discriminative power. Most of these tests were developed for use in college student and high school populations.

It was observed, on hand scoring the protocols of the D.R.T. that many of the answer sheets were incomplete. It seemed possible in view of this, that the test might be somewhat speeded for the sample tested. This might account, at least in part, for the high observed internal consistency coefficients. On further inspection of the protocols it was found, in fact, that whereas

85% of the sample had completed the test up to item 36, only 60% managed to complete all 45 items.

In order to test whether this circumstance was responsible for inflating the reliability estimates, an arbitrary cut-off was taken at item 36 and separate scores were recorded for each subject on the shortened 36 item version of the test. It was argued that if the scores on the reduced (unspeeded) version of the test correlated poorly with scores on the total test, a time considera-tion could be interfering with the performance of certain subjects. This would imply that a significant proportion of the reliable variance of the test is due to a speed, rather than a power compo-nent. Fortunately this did not prove to be the case. The two versions of the test were in fact highly correlated (0,96) indicating that the better subjects were consistently better and also managed to complete more items in the allotted time. The re-liability estimates can therefore be assumed to be realistic assessments of the power component of the test. It seems likely, considering that the items in the test were ordered em-pirically in ascending order of difficulty, that the difficulty levels of items beyond number 36 in the test were in excess of the latent abilities of many subjects. In other words, subjects found each subsequent item a little more difficult than the immediately pre-ceding one until, in the case of about 40% of the sample (i.e. those subjects who did not complete the test), the items became too difficult to solve. This impression was confirmed in an examination of the error frequencies in the protocols of subjects who failed to complete the test. The number of errors made on attempted items increased progressively toward the end of the test.

The attainment of this circumstance fulfils a major aim of the study, namely to develop an instrument capable of measuring deductive reasoning ability reliably in an intellectually sophisti-cated population.

The monotonically increasing distribution of the difficulty values can be seen in Table 3. The difficulty coefficient for each item indicates the proportion of subjects that answered the item correctly. The column p_j in Table 3 denotes the difficulty values. The column n_j indicates the actual number of subjects attempting each item. The item standard deviations (S_j) and item reliabilities ($r_{xj}S_j$) are also shown in Table 3.

It will be seen from an inspection of the table that a number of items at the beginning of the test proved too easy for the CSIR sample in general. These items accordingly had negligible variances and consequently contributed little toward the test variance, and hence the reliability of the total test. Disregarding these few items, however, it can be seen that the reliabilities of individual items were generally good.

An Item Response Evaluation analysis was also performed. Since the program provides detailed information about the characteristics of each individual item and of each distractor in each item, the results are too bulky to tabulate or to present in a discussion of this nature. In summary of the findings from this analysis, however, it may be reported that with the exception of those items at the beginning of the test that proved too easy, all items functioned extremely well. With few exceptions, the distractors appeared to have been well-chosen and discriminated with appreciable reliability among subjects. A comparison between the item characteristics obtained in this analysis and those obtained in the analysis of test X, Y and Z (Verster,[41)] 1972) seems to suggest that the difficulty level of the test and the complexity of the task, were better suited to the slightly more advanced CSIR research population. This is indeed fortunate since the final selection instrument is to be used in this particular population.

TABLE 3.

DIFFICULTIES, STANDARD DEVIATIONS AND RELIABILITIES OF 45 D.R.T. ITEMS

ITEM	n_j	p_j	s_j	$r_{xj}s_j$
1	158	0,99	0,11	-0,00
2	154	0,96	0,19	0,01
3	156	0,98	0,16	0,02
4	150	0,94	0,24	0,07
5	146	0,91	0,28	0,09
6	144	0,90	0,30	0,11
7	146	0,91	0,28	0,03
8	141	0,88	0,32	0,11
9	135	0,84	0,36	0,11
10	129	0,81	0,40	0,15
11	141	0,88	0,32	0,09
12	130	0,81	0,39	0,19
13	107	0,67	0,47	0,15
14	130	0,81	0,39	0,21
15	147	0,92	0,27	0,09
16	116	0,73	0,45	0,22
17	103	0,64	0,48	0,13
18	125	0,78	0,41	0,12
19	133	0,83	0,38	0,10
20	128	0,80	0,40	0,22
21	125	0,78	0,41	0,15
22	119	0,74	0,44	0,16
23	135	0,84	0,36	0,16
24	118	0,74	0,44	0,14
25	131	0,82	0,39	0,16
26	89	0,56	0,50	0,17
27	117	0,73	0,44	0,25
28	105	0,66	0,48	0,24
29	89	0,56	0,50	0,28
30	92	0,58	0,49	0,19
31	91	0,57	0,50	0,25
32	112	0,70	0,46	0,24
33	90	0,56	0,50	0,25
34	103	0,64	0,48	0,25
35	82	0,51	0,50	0,27
36	69	0,43	0,50	0,27
37	100	0,63	0,48	0,26
38	81	0,51	0,50	0,30
39	99	0,62	0,49	0,31
40	74	0,46	0,50	0,24
41	68	0,43	0,49	0,32
42	79	0,49	0,50	0,32
43	51	0,32	0,47	0,26
44	49	0,31	0,46	0,19
45	93	0,58	0,49	0,28

3.2.1.2. Construction of the final version of the Deductive Reasoning Test

Certain explicitly formulated criteria guided the selection of items included in the final form of the Deductive Reasoning Test. Firstly, as in the compilation of the 45 item version of the test, an attempt was made to attain an empirically determined platykurtic distribution of item difficulties. This type of difficulty distribution has the advantage of ensuring a maximum range of effective, reliable discrimination among individual test scores. Examination of the p_j column in Table 3 reveals that, even if the first set of six or seven very easy items is disregarded, the distribution of difficulty values is still slightly skewed in the negative direction. There is thus a slight overrepresentation of easy items in the available item pool. In this regard, it should be borne in mind that the observed difficulty value of an item is not a fixed property of the item, but a function of the mean level of the experimental population's latent ability to cope with the mental task set by the item. It is likely therefore that when the final selection test is applied to the general population of CSIR applicants, there will be a downward shift in the mean difficulty level of items since the experimental sample represents a pre-selected sample of successful candidates. If the success of an applicant is partially dependent on his scores on the cognitive tests in the selection battery, which is the case in the present situation, then it can be argued that the mean ability of successful applicants will be higher than the mean ability of the a priori population. The expected general shift in item difficulties when the final test is applied to the applicant population should cause the slight skewness in the present difficulty distribution to disappear altogether. But the relative difficulty values of items should nevertheless remain more or less constant in the new population since the items within the test are ordered according to difficulty in terms of empirically derived values from separate studies conducted in two different populations.

For practical reasons, a test with a 40 minute time limit was considered desirable. In terms of the average time per item taken by the CSIR sample, this meant producing a final version containing 36 items.

On the basis of the item reliabilities ($r_{xj}S_j$) in Table 3, 9 items from the 45 item test were omitted. These chiefly constituted very easy items from the beginning of the test.

The ordering of items in the final version of the Deductive Reasoning Test is shown in Table 4. The difficulty values are provided in the same table. The test has been made available in this form for use in the selection of CSIR applicants.

It should be borne in mind that the 45 item version of the test was used in all the analyses of the test battery that are discussed later.

3.2.2. Test Characteristics of the Gevorderde Woordeskattoets (G.W.S.)

The 50 item version of this test had a mean of 24,2 and a standard deviation of 6,87. If anything, the test was slightly difficult for the population tested. The reliability of the instrument for the sample of 72 Afrikaans-speaking subjects, estimated with KR_{20} was 0,82.

The results of an item analysis performed on the test are shown in Table 5.

By an iterative procedure, successive items were omitted from the test until the reliability of the reduced instrument reached a maximum of 0,90. At this point 30 items were left in the test. A distractor analysis confirmed that these 30 items contained no deficiencies in the distractors. The test was thus judged satisfactory and the shortened 30 item version could be included in the test battery for the purpose of further analyses.

Table 4.

DEDUCTIVE REASONING TEST (36 ITEMS)

Item No.	Original Position	Difficulty	Dimensions
1	6	0,900	$C_3\ F_1\ M_2$
2	8	0,881	$C_2\ F_1\ M_6$
3	9	0,844	$C_2\ F_3\ M_8$
4	23	0,844	$C_2\ F_1\ M_7$
5	25	0,819	$C_2\ F_2\ M_4$
6	12	0,813	$C_1\ F_3\ M_2$
7	14	0,813	$C_3\ F_3\ M_5$
8	10	0,806	$C_2\ F_2\ M_6$
9	20	0,800	$C_2\ F_2\ M_7$
10	18	0,781	$C_3\ F_4\ M_5$
11	21	0,781	$C_2\ F_4\ M_3$
12	22	0,744	$C_3\ F_3\ M_7$
13	24	0,737	$C_1\ F_2\ M_4$
14	27	0,731	$C_2\ F_3\ M_5$
15	16	0,725	$C_3\ F_1\ M_7$
16	32	0,700	$C_1\ F_4\ M_5$
17	13	0,669	$C_1\ F_3\ M_5$
18	28	0,656	$C_3\ F_3\ M_8$
19	17	0,644	$C_3\ F_2\ M_4$
20	34	0,644	$C_2\ F_4\ M_5$
21	37	0,625	$C_2\ F_4\ M_5$
22	39	0,619	$C_1\ F_2\ M_7$
23	45	0,581	$C_2\ F_1\ M_2$
24	30	0,575	$C_2\ F_4\ M_1$
25	26	0,556	$C_2\ F_3\ M_2$
26	33	0,563	$C_3\ F_3\ M_1$
27	31	0,569	$C_3\ F_2\ M_7$
28	29	0,556	$C_1\ F_4\ M_5$
29	35	0,512	$C_3\ F_2\ M_7$
30	40	0,462	$C_2\ F_3\ M_6$
31	42	0,494	$C_1\ F_2\ M_7$
32	38	0,506	$C_1\ F_4\ M_7$
33	36	0,431	$C_2\ F_3\ M_6$
34	41	0,425	$C_1\ F_3\ M_7$
35	43	0,319	$C_3\ F_3\ M_6$
36	44	0,306	$C_2\ F_3\ M_1$

3.2.3. Problems with the Reasoning Ability Test (R.A.T.)

Due to a rather restricted test variance, the R.A.T. yielded a reliability of 0,50 using KR_{20}. The test was not considered suitable for inclusion in the main analysis in this form. But in preference to omitting one of the markers for the postulated factor of deduction, an attempt was made to salvage the instrument.

An item analysis revealed that in addition to the test's poor discriminative power, the low reliability estimate was caused by certain items correlating poorly, or not at all, with the total test score. Since KR_{20} is essentially an estimate of internal consistency, a deflated estimate of this coefficient could be expected when dealing with a heterogeneous test. By rejecting items with poor item-test correlations, the reliability was increased to 0,57 based on a sample of 160 cases. Although a coefficient of this magnitude would not justify the use of a test for any predictive purpose, it was considered sufficiently high to warrant the inclusion of the R.A.T. in the analysis of the battery. The modified Reasoning Ability Test consisted of 18 items.

3.2.4. Problems with the Locations Test

In an earlier section it was observed that the Locations Test, which consists of 28 items, proved to be too highly speeded for the CSIR sample. Only 18% of the sample managed to complete the test in the allocated time. This could, at least in part, be attributed to the fact that the test was administered first in the battery and many of the subjects were unaccustomed to the speeded working conditions. It should be stressed, in this regard, that the prescribed time limit for the test, as with the other E.T.S. tests in the battery, was not adhered to. The time limits suggested by French et al.[42)] (1963) in their Kit were considered too stringent for use in the slightly slower paced

Table 5.

DIFFICULTIES, STANDARD DEVIATIONS AND RELIABILITIES OF 50 GEVORDERDE WOORDESKATTOETSITEMS

ITEM	n_j	p_j	s_j	$r_{xj}s_j$
1	69	0,96	0,20	-0,05
2	28	0,39	0,49	0,09
3	10	0,14	0,35	0,06
4	66	0,92	0,28	0,09
5	48	0,67	0,47	0,16
6	16	0,22	0,42	0,23
7	34	0,47	0,50	0,14
8	67	0,93	0,25	0,04
9	34	0,47	0,50	0,30
10	6	0,08	0,28	0,10
11	56	0,78	0,42	0,17
12	39	0,54	0,50	0,08
13	57	0,79	0,41	0,02
14	14	0,19	0,40	0,24
15	31	0,43	0,50	0,30
16	69	0,96	0,20	0,01
17	20	0,28	0,45	0,20
18	59	0,82	0,39	0,09
19	6	0,08	0,28	0,07
20	17	0,24	0,43	0,15
21	68	0,94	0,23	0,03
22	71	0,99	0,12	0,01
23	52	0,72	0,45	0,22
24	23	0,32	0,47	0,14
25	51	0,71	0,46	0,06
26	35	0,49	0,50	0,02
27	48	0,67	0,47	0,28
28	59	0,82	0,39	0,05
29	13	0,18	0,39	0,06
30	41	0,57	0,50	0,30
31	7	0,10	0,30	0,08
32	67	0,93	0,25	0,10
33	56	0,78	0,42	0,17
34	69	0,96	0,20	0,02
35	63	0,88	0,33	0,17
36	21	0,29	0,46	0,20
37	39	0,54	0,50	0,19
38	22	0,31	0,46	0,14
39	28	0,39	0,49	0,27
40	15	0,21	0,41	0,24
41	42	0,58	0,49	0,21
42	39	0,54	0,50	0,19
43	30	0,42	0,49	0,24
44	25	0,35	0,48	0,28
45	68	0,94	0,23	0,05
46	7	0,10	0,30	0,13
47	42	0,58	0,50	0,18
48	49	0,68	0,47	0,22
49	9	0,13	0,33	0,06
50	11	0,15	0,36	0,01

South African cultural context. Extensions to the suggested time limits were therefore permitted in order to exclude the possibility of contaminating the desired sources of variance in the tests with a speed component. Yet, despite this precaution, Locations proved too highly speeded. An attempt to partial out the undesired speed variance in the test was therefore indicated.

An examination of the test protocols revealed that about 80% of the sample managed to complete the test up to and including item 14. This represents the first half of the test. Yet, as indicated above, only 18% completed the entire test. The reliability of the whole test estimated with KR_{20} was 0,54. This coefficient was increased to 0,67 when only the first 14 items in the test were considered. The difference between the two estimates is attributed to the contaminating effects of a speed component which placed a ceiling on the spread of scores over the whole test. The reliability based on the first half of the test, which was relatively unspeeded, is regarded as a close approximation to the true internal consistency of the test in terms of a power component. The Spearman Brown extension formula indicates that the internal consistency would increase to the order of 0,8 if the number of items were doubled ($K = 28$), while the same average time per item is permitted. In the light of this finding it was considered advisable to use only the first half of the test in the analyses. The fact that a correlation of 0,79 was obtained between scores on the first 14 items and the full test, was taken as further justification for adopting this strategy. This coefficient was interpreted as implying that it was unlikely that a significant change in the rank order of subjects, in terms of the power component of the test, would be brought about. It is possible though, that by ignoring the second half of the test, the rank order of subjects, who attempted more than 14 items, would be slightly altered in terms of a speed component. The investigation, however, is concerned with the nature and order of the inter-correlations between the power components in the tests in the battery. None of the other tests in the battery appears to have been too greatly contaminated with speed variance.

4. TEST STATISTICS OF THE BATTERY

Before proceeding to a discussion of the correlational analyses performed on the test battery, attention will be focussed on certain pertinent descriptive statistics of the tests. Prior to calculating these statistics, the sample was divided into separate English- and Afrikaans-speaking sub-samples on the basis of home language. The English sample comprised 88 subjects while the Afrikaans sample size was 72. The author considered it necessary to partition the sample in terms of home language in order to satisfy the assumption that the two language groups were statistically representative of the same overall population in terms of their test performance. If, for example, significant differences between the sample means are found, it would have to be concluded that the two language groups represent different populations. In this event, further statistical treatment of the data based on the assumption that the scores were obtained from a single, homogeneous population would be unjustified. From the results that follow it was concluded unequivocally, that the differences in test performance between the two samples indicate that they are from different populations. Parallel analyses on the separate samples were therefore performed.

The test means, standard deviations, coefficients of skewness and kurtosis and reliabilities calculated separately for the two language groups appear in Tables 6 and 7.

An examination of the means and skewnesses reveals that the tests were generally of an appropriate level of difficulty in both samples. In most cases the slight negative skewnesses are sufficiently small to be discounted. Cube Comparison and Advanced Vocabulary were perhaps too easy in the English sample. Plots were drawn of the frequency distributions of all tests. These confirm the indices of skewness and kurtosis. Only Cube Comparison was found to have bi-modal distribution of scores in the Afrikaans sample. No acceptable explanation for this unexpected distribution could be found.

An inspection of the standard deviations suggests that a better spread of test scores would have been desirable in several cases. In particular, the tests drawn from the E.T.S. reference kit appear to have lacked discriminative power. It should be borne in mind, in evaluating these statistics, that both samples were relatively small and very homogeneous. It is therefore not altogether surprising that some of the tests failed to discriminate as well as they might in a more heterogeneous population. It is gratifying to note that both the Deductive Reasoning Test and the Gevorderde Woordeskattoets that were constructed specifically for use in these populations, were able to discriminate well. In general though, the mesokurtic frequency distributions have caused the reliability estimates to be depressed. Reliabilities were estimated by a convenient approximation to Kuder-Richardson formula 20. This formula has been shown to provide very close approximations to KR_{20}, although if anything the estimates are usually smaller. (Cf. Saupe,[43] 1961). The formula is:

$$r_{tt} = 1 - \frac{0,19K}{\sigma^2}$$

where K = number of items in the test

σ^2 = variance of the test

It is interesting to note the effect that test length has on the reliability estimates. For example Figure Classification, which consists of 224 items, yielded reliabilities of 0,95 and 0,96. These are considered to be unrealistically high estimates. They have probably been inflated due to the large speed component in

TABLE 6

ENGLISH SAMPLE N = 88

MEANS, STANDARD DEVIATIONS, COEFFICIENTS OF SKEWNESS AND KURTOSIS AND RELIABILITIES OF FOURTEEN ABILITY TESTS

	Test	Mean	Standard Deviation	Skewness	Kurtosis	Reliability
1	Locations	8,66	2,41	-0,59	0,70	0,54
2	D.R.T.	33,19	8,14	-0,76	-0,05	0,87
3	Card Rotation	21,14	4,53	-0,59	-0,40	0,74
4	H.L. Vocabulary	21,53	6,77	-0,19	-0,43	0,83
5	Letter Sets	23,22	3,61	-0,29	-0,02	0,56
6	Inference	16,90	2,79	-0,02	-0,92	0,51
7	Cube Comparison	33,66	6,21	-1,14	1,18	0,79
8	H.L. Reading Comp.	14,22	2,93	-0,87	2,10	0,66
9	Figure Class.	140,55	33,74	-0,31	-0,26	0,96
10	Blox	35,17	3,97	-0,10	-0,22	0,46
11	Reasoning Ability	12,34	2,49	-0,57	0,38	0,45
12	Pattern Comp.	18,65	5,79	-0,04	0,59	0,83
13	Figure Series	21,67	5,57	-0,67	-0,45	0,82
14	Adv. Vocabulary	29,00	5,50	-1,03	0,68	0,82

TABLE 7

AFRIKAANS SAMPLE N = 72

MEANS, STANDARD DEVIATIONS, COEFFICIENTS OF SKEWNESS AND KURTOSIS AND RELIABILITIES OF FOURTEEN ABILITY TESTS

Test		Mean	Standard Deviation	Skewness	Kurtosis	Reliability
1	Locations	7,75	2,49	-0,28	-0,07	0,57
2	D.R.T.	30,07	7,80	-0,44	-0,13	0,86
3	Card Rotation	20,28	4,91	-0,21	-0,14	0,68
4	H.L. Vocabulary	27,86	5,30	-0,44	0,02	0,73
5	Letter Sets	21,85	4,63	-0,57	0,16	0,73
6	Inference	12,19	2,85	-0,52	0,45	0,53
7	Cube Comparison	30,60	7,39	-0,63	0,37	0,85
8	H.L. Reading Comp.	13,00	2,87	-0,39	0,92	0,64
9	Figure Class.	129,49	30,10	0,39	-0,20	0,95
10	Blox	34,06	5,09	-0,12	0,77	0,67
11	Reasoning Ability	11,50	3,01	-0,85	0,17	0,62
12	Pattern Comp.	17,46	6,26	-0,12	0,47	0,81
13	Figure Series	20,22	5,03	-0,40	-0,30	0,73
14	Adv. Vocabulary	26,60	6,69	0,10	-0,55	0,87

the test. Tests with relatively few items, on the other hand, were found to yield low reliabilities. Although a number of the reliability estimates are very low, notably in the case of Reasoning Ability, Blox and Inference in the English sample and Inference in the Afrikaans sample, it was decided to retain these tests in the analyses.

It is contended that the low reliability indices reflect a restricted range of test variance rather than a lack of internal consistency. Despite the poor reliabilities that were obtained, the tests are regarded as relatively "pure" ability measures, since they have been shown to yield high internal consistency coefficients in larger, more heterogeneous samples. The problem in the present study appears to be that the tests were less suited to the task of discriminating effectively in the very homogeneous, pre-selected samples. Considerable caution should therefore be exercised when evaluating the implications of these results. It is contended that the poor reliability estimates will not greatly affect the stability of the correlations between tests since few of the tests are likely to be non-homogeneous. Moreover, it should be observed that the correlation between two variables cannot exceed the square root of the product of their reliabilities. Despite this limitation placed on the possible magnitude of a correlation, it is theoretically possible for a test with a reliability of 0,50 to correlate in the order of 0,60 with a test having a reliability of 0,80. In view of these considerations, it seemed worthwhile to retain all the tests in the battery for the purpose of the correlational analyses. The results obtained from these analyses should, however, be interpreted with great caution. The findings that emerge from the analyses to be described in the following sections will be strengthened if they make good sense on psychological grounds and can be reconciled with known theory.

A comparison between the test means in Tables 6 and 7 reveals that, disregarding the language tests, the means of the remaining 11 tests are higher in the English sample. Since the three language tests are not directly comparable across samples, these should be ignored for the purpose of the present discussion. The consistent trend in mean test score differences in favour of the English-speaking sample was not entirely unexpected. Differences in the same direction have been observed before in unpublished NIPR studies. They are generally ascribed to a cultural pre-selection in terms of occupational preferences.

In order to determine the statistical significance of the observed differences, t-tests were calculated. The results are summarized in Table 8 on Page 40.

It can be seen that 6 of the 11 differences are significant at the 5% level of confidence or better. The other differences, although not statistically significant, are without exception in the same direction, thus giving strength to the general trend observed.

In terms of their mean performances on a wide variety of cognitive tests, it can consequently be asserted that the two samples concerned derive from two separate populations. This finding is considered most interesting and valuable from a theoretical point of view. The practical implications for the immediate study are that under no circumstances can the test scores of the two groups be pooled in subsequent analyses. Separate, but parallel analyses will be undertaken.

TABLE 8

COMPARISON BETWEEN THE ENGLISH AND AFRIKAANS SAMPLE MEANS

	Test	Difference	t-value	(P) Significance
1	Locations	0,91	2,33	*
2	D.R.T.	3,12	2,47	*
3	Card Rotation	0,86	1,14	
4	Letter Sets	1,37	2,07	*
5	Inference	4,71	11,94	***
6	Cube Comparison	3,06	2,78	**
7	Figure Class.	11,06	2,19	*
8	Blox	1,11	1,52	
9	Reasoning Ability	0,84	1,90	
10	Pattern Comp.	1,19	1,24	
11	Figure Series	1,45	1,73	

d.f = 158

* = Significant at 5% level of confidence

** = Significant at 1% level of confidence

*** = Significant at 0,1% level of confidence

It would be most tempting, at this point, to attempt an explanation of the observed differences in mean performance between the two samples. But in terms of the perticular experimental design employed in the present study it is not possible to offer sound explanations for these differences. Without additional evidence the results cannot be interpreted as demonstrating a general intellectual superiority of one language group over the other. Indeed, it seems very unlikely that this is the case. The observed differences may simply be due to cultural differences in test taking attitudes or to differences in cognitive style. A host of cultural factors including informal differences in the way the formal educational programs are dealt with in the two sub-cultures could also lie at the root of the differences. It should be borne in mind that this study was not undertaken with a view to examining mean differences in test performance between the two language groups and certain very important sampling variables in this respect have consequently been left uncontrolled. For example, it is possible that a greater number of Afrikaans-speaking subjects received their formal schooling in the rural and country districts, where the quality of teaching is often not as good as that in the cities and urban areas. This factor alone, could be responsible for the observed differences. In addition sampling should have considered occupational preferences. An unpublished study undertaken at the NIPR (Steyn, 1969) has shown that the two language-cultural groups differ along important and highly relevant personality dimensions such as rigidity (over regimentation), which could account for differences in stylistic approach to cognitive tests. Further research into this very interesting problem, in the form of carefully controlled cross-cultural studies is strongly recommended.

5. INTERCORRELATIONS AND COVARIANCES

The separate intercorrelation matrices for the English- and Afrikaans-speaking samples respectively are shown in Tables 9 and 10. Only the lower halves of the matrices are represented. A visual comparison of these two tables is enough to demonstrate how very different the intercorrelation patterns are in the two groups. It seems that the average order of the intercorrelations in the Afrikaans sample is greater than that in the English sample. This suggests that the resultant factorial structure in the English sample might point to a more clearly differentiated intellectual structure. It will be seen presently whether this is, indeed, the case.

The differences between individual correlations in the two samples are at times striking and certainly difficult to account for. Consider for example the intercorrelations between the D.R.T. and the first test in the battery, Locations. In the English sample these tests yield a Pearson product-moment correlation coefficient of 0,29. In the Afrikaans sample the value is 0,56. Undoubtedly this difference would be statistically significant. Similarly the two vocabulary tests intercorrelate in the order of 0,81 in the English sample and 0,66 in the Afrikaans sample; again a striking difference.

The picture is no less confusing when the intercorrelations are weighted by the standard deviations in the form of covariances. The covariance between two variables is given by the product of their intercorrelation coefficient and their respective standard deviations.

The covariance matrices obtained for the two samples are shown in Tables 11 and 12.

TABLE 9

ENGLISH SAMPLE N = 88

INTERCORRELATION MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Locations	1,00													
2 D.R.T.	0,29	1,00												
3 Card Rotation	0,31	0,33	1,00											
4 H.L. Vocabulary	0,28	0,25	-0,06	1,00										
5 Letter Sets	0,32	0,40	0,37	0,13	1,00									
6 Inference	0,24	0,21	0,11	0,31	0,26	1,00								
7 Cube Comparison	0,19	0,26	0,45	-0,13	0,15	0,12	1,00							
8 H.L. Reading Comp.	0,41	0,28	-0,01	0,40	0,10	0,27	-0,12	1,00						
9 Figure Class.	0,21	0,59	0,40	0,21	0,37	0,15	0,39	0,14	1,00					
10 Blox	0,28	0,34	0,44	0,06	0,42	0,02	0,43	0,09	0,46	1,00				
11 Reasoning Ability	0,32	0,36	0,13	0,24	0,38	0,23	0,17	0,13	0,24	0,25	1,00			
12 Pattern Comp.	0,41	0,30	0,30	-0,00	0,39	0,22	0,24	0,29	0,34	0,43	0,31	1,00		
13 Figure Series	0,18	0,27	0,35	0,07	0,26	0,14	0,17	0,13	0,36	0,30	0,04	0,56	1,00	
14 Adv. Vocabulary	0,26	0,24	0,04	0,81	0,11	0,31	-0,13	0,39	0,19	0,04	0,22	-0,02	0,06	1,00

TABLE 10

AFRIKAANS SAMPLE N = 72

INTERCORRELATION MATRIX

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Locations	1,00													
2	D.R.T.	0,56	1,00												
3	Card Rotation	0,50	0,39	1,00											
4	H.L. Vocabulary	0,22	0,23	0,06	1,00										
5	Letter Sets	0,52	0,43	0,37	0,41	1,00									
6	Inference	0,46	0,51	0,19	0,36	0,50	1,00								
7	Cube Comparison	0,50	0,45	0,49	0,33	0,47	0,48	1,00							
8	H.L. Reading Comp.	0,35	0,43	0,22	0,26	0,30	0,50	0,38	1,00						
9	Figure Class.	0,62	0,48	0,26	0,12	0,54	0,35	0,46	0,34	1,00					
10	Blox	0,49	0,31	0,55	0,28	0,41	0,28	0,50	0,25	0,31	1,00				
11	Reasoning Ability	0,53	0,48	0,37	0,26	0,47	0,41	0,46	0,43	0,46	0,23	1,00			
12	Pattern Comp.	0,43	0,43	0,39	0,22	0,55	0,49	0,55	0,46	0,54	0,56	0,51	1,00		
13	Figure Series	0,41	0,24	0,30	0,15	0,48	0,37	0,37	0,43	0,42	0,39	0,51	0,65	1,00	
14	Adv. Vocabulary	0,08	0,09	0,01	0,66	0,30	0,38	0,25	0,21	0,10	0,08	0,14	0,12	0,09	1,00

TABLE 11

ENGLISH SAMPLE N = 88

CONVARIANCE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 LOCATION	505,78													
2 D.R.T.	501,80	5767,75												
3 CARD ROTATION	293,09	1070,68	1782,37											
4 H.L. VOCABULARY	395,02	1206,92	-153,41	3983,90										
5 LETTER SETS	243,82	1015,19	531,27	283,32	1135,46									
6 INFERENCE	99,93	298,74	87,23	360,81	160,05	336,08								
7 CUBE COMPARISON	251,09	1132,19	1102,37	-462,91	285,31	132,73	3356,38							
8 H.L. READING COMP.	213,48	493,33	75,59	580,86	79,68	115,95	-158,09	538,90						
9 FIGURE CLASS.	1467,38	14172,80	5318,50	4214,38	3922,13	844,94	7124,50	1008,69	99064,00					
10 BLOX	231,12	951,13	685,96	148,00	521,63	11,54	928,50	-77,24	5334,88	1382,50				
11 REASONING ABILITY	166,23	636,21	130,91	344,98	297,18	97,07	231,91	72,52	1764,69	211,89	537,77			
12 PATTERN COMP.	502,43	1249,99	675,23	-441,41	716,05	218,83	752,73	358,70	5831,94	855,29	390,57	2918,08		
13 FIGURE SERIES	214,12	1063,61	768,96	225,49	462,59	134,04	509,46	154,26	5807,88	572,00	51,89	1574,79	2695,45	
14 ADV. VOCABULARY	294,00	947,00	79,00	2614,00	186,00	290,00	-380,00	460,00	3027,00	71,00	262,00	-58,00	156,00	2628,00

TABLE 12

AFRIKAANS SAMPLE N = 88

COVARIANCE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 LOCATIONS	439,50													
2 D.R.T.	773,25	4322,66												
3 CARD ROTATION	432,00	1048,61	1708,45											
4 H.L. VOCABULARY	203,50	664,70	109,78	1994,61										
5 LETTER SETS	421,25	1098,77	395,06	707,47	1521,32									
6 INFERENCE	230,50	801,03	192,11	381,95	471,14	575,28								
7 CUBE COMPARISON	651,75	1856,06	1268,06	925,97	1135,57	714,64	3877,38							
8 H.L. READING COMP.	150,00	583,00	188,00	235,00	242,00	247,00	481,00	424,00						
9 FIGURE CLASS.	3310,75	7996,69	2771,31	1415,94	5372,38	2148,25	7207,25	1756,00	64337,00					
10 BLOX	437,00	863,75	977,89	529,56	678,61	284,22	1324,63	222,00	3363,25	1841,81				
11 REASONING ABILITY	282,00	803,50	389,00	295,00	464,50	252,00	727,50	223,00	2960,50	255,00	644,00			
12 PATTERN COMP.	477,25	1484,71	859,84	518,59	1138,04	614,59	1805,29	502,00	7198,00	1257,17	688,50	2785,88		
13 FIGURE SERIES	363,00	677,89	527,56	291,22	786,45	375,89	988,45	377,00	4547,25	703,11	543,00	1454,67	1794,45	
14 ADV. VOCABULARY	95,75	324,02	16,06	1650,97	654,57	518,64	887,32	241,00	1439,25	181,61	201,50	365,29	219,45	3177,32

In order to confirm whether the covariance matrices obtained for the two language groups were significantly different, a Box-Wilks Test was performed. A lamda (λ) coefficient of 147,709 was obtained. With 105 degrees of freedom, this value is statistically significant at the 99,9% level of confidence.

On the basis of these results the two samples indisputably represent different populations. Separate factor analyses as well as separate multi-dimensional scaling analyses on the two sets of data were therefore performed.

6. FACTOR ANALYSIS

6.1. The Criterion of Simple Structure

Thurstone[44)] (1947, p. 319) stated "the principles of simple structure are fundamental in making factor analysis a scientific method rather than merely a method of statistical condensation, and it is therefore one of the central themes in factorial theory." In its essence, the principle of simple structure assumes that, in a given battery of tests to be factored, the complexity of each independent test is less than the complexity (r) of the battery as a whole. If this is indeed the case, there will be one or more blank cells in each row of the factor pattern. The resultant factor pattern is then considered as evidence of an underlying simple structure in which not every mental process or parameter is involved in every test measurement. The simple structure will be orthogonal if the factors are uncorrelated in the experimental population, while if they are correlated, it will be indicative of an oblique simple structure. The principle of simple structure does not leave the important question of where to locate the factor reference axes to numerical chance. It implies that some degree of rotation, either to an orthogonal, or to an oblique position is necessary to attain the desired simple structure. By attaining a simple structure, the psychological meaningfulness of the factors will be facilitated. In the analyses special statistical criteria were used to indicate the degree of closeness to simple structure that was obtained.

The decision rule that was followed in estimating the number of factors to extract was Kaiser's[45] (1970) Little Jiffy 2 criterion. According to this criterion, the number of factors is equal to the number of eigen values of the weighted correlation matrix greater than or equal to the specified cut-off value proposed by the criterion statistic. The number of factors for extraction was specified as three in both the English and the Afrikaans samples. Since four factors were expected to account for the intercorrelations in the test battery in terms of the design of the experiment, it was decided that separate analyses should be undertaken extracting three and four factors alternatively. The program that was used in the analyses made this step an easy one since rotations can be repeated for a different number of factors without repeating the entire factor analysis. Considerations of psychological interpretability would ultimately determine which solution to accept. Kaiser's[46] (1965, p. 43) rescaled simplicity criterion was used to indicate which solution resulted in the closest approximation to simple structure in the case of the orthogonal (varimax) rotations. In the case of the oblique (direct quartimin) rotations the rescaled criterion of Jerrich and Sampson[47] (1966) was used.

A non-iterating Jöreskog[48] (1963) factor analysis procedure was employed. In terms of this procedure, the factor that accounts for the greatest proportion of variance is printed first. Successively printed factors account for progressively less of the variance in the test battery. The number of factors to be extracted was specified as three and four alternatively in the case of both language groups.

The four-factor solution, whether rotated to an orthogonal or to an oblique position, proved unsuccessful in both analyses. In terms of the specified criteria the three factor solutions gave better approximations to simple structure in both cases. It appears that the rank of the intercorrelation matrices in both sets of data is definitely less than four. The Kaiser "Little Jiffy 2" decision rule thus seems justified in suggesting the extraction of three factors in each sample. When more than three factors are extracted, some form of factor fission appears to take place causing a pair of variables that are highly correlated with one another to split off and form a separate "artificial" factor which defies meaningful interpretation.

Three factors were therefore specified for extraction in both the English and Afrikaans data. The orthogonal rotations were obtained by means of the Varimax procedure. The rotated, orthogonal three factor solutions are presented in Tables 13 and 14.

The oblique solutions, using the direct quartimin method of rotation, are given in Tables 15 and 17. The factor correlation matrices are given in Tables 16 and 18 respectively.

TABLE 13

ENGLISH SAMPLE N = 88

THREE FACTORS (VARIMAX ROTATION)

	Test	I	II	III	h^2
1	Locations	0,24	0,29	0,48	0,37
2	D.R.T.	0,54	0,30	0,24	0,44
3	Card Rotation	0,62	-0,04	0,18	0,42
4	H.L. Vocabulary	0,02	0,87	0,04	0,75
5	Letter Sets	0,47	0,13	0,34	0,36
6	Inference	0,10	0,35	0,28	0,21
7	Cube Comparison	0,60	-0,15	0,04	0,39
8	H.L. Reading Comp.	0,13	0,45	0,47	0,44
9	Figure Class.	0,67	0,21	0,16	0,52
10	Blox	0,67	0,01	0,18	0,48
11	Reasoning Ability	0,30	0,26	0,25	0,22
12	Pattern Comp.	0,35	-0,04	0,71	0,63
13	Figure Series	0,34	0,[illegible]1	0,48	0,34
14	Adv. Vocabulary	0,02	0,86	0,02	0,74

TABLE 14

AFRIKAANS SAMPLE N = 72

THREE FACTORS (VARIMAX ROTATION)

	Test	I	II	III	h^2
1	Locations	0,63	0,42	0,05	0,57
2	D.R.T.	0,61	0,23	0,11	0,44
3	Card Rotation	0,27	0,66	-0,04	0,51
4	H.L. Vocabulary	0,14	0,14	0,75	0,61
5	Letter Sets	0,56	0,33	0,32	0,53
6	Inference	0,59	0,11	0,41	0,53
7	Cube Comparison	0,47	0,48	0,27	0,53
8	H.L. Reading Comp.	0,55	0,11	0,22	0,36
9	Figure Class.	0,69	0,22	0,01	0,53
10	Blox	0,23	0,73	0,13	0,60
11	Reasoning Ability	0,66	0,20	0,12	0,49
12	Pattern Comp.	0,62	0,45	0,12	0,60
13	Figure Series	0,57	0,31	0,06	0,43
14	Adv. Vocabulary	0,09	-0,02	0,76	0,58

TABLE 15

ENGLISH SAMPLE N = 88

THREE FACTORS (DIRECT QUARTIMIN ROTATION)

	Test	I	II	III	h^2
1	Locations	0,12	0,16	0,48	0,37
2	D.R.T.	0,52	0,26	0,12	0,44
3	Card Rotation	0,62	-0,06	0,06	0,42
4	H.L. Vocabulary	0,03	0,88	-0,03	0,75
5	Letter Sets	0,41	0,05	0,27	0,36
6	Inference	0,03	0,27	0,27	0,21
7	Cube Comparison	0,64	-0,14	-0,09	0,39
8	H.L. Reading Comp.	-0,27	0,31	0,54	0,44
9	Figure Class.	0,68	-0,21	0,01	0,52
10	Blox	0,68	-0,01	0,04	0,48
11	Reasoning Ability	0,26	0,21	0,19	0,22
12	Pattern Comp.	0,16	-0,25	0,75	0,63
13	Figure Series	0,22	-0,13	0,48	0,34
14	Adv. Vocabulary	0,04	0,88	-0,06	0,74

TABLE 16

ENGLISH SAMPLE N = 88

FACTOR CORRELATION MATRIX

	I	II	III
I	1,00	-	-
II	0,03	1,00	-
III	0,44	0,35	1,00

TABLE 17

AFRIKAANS SAMPLE N = 72
THREE FACTORS (DIRECT QUARTIMIN ROTATION)

	Test	I	II	III	h^2
1	Locations	0,61	0,24	-0,08	0,57
2	D.R.T.	0,66	0,02	-0,02	0,44
3	Card Rotation	0,08	0,67	-0,08	0,51
4	H.L. Vocabulary	0,00	0,12	0,76	0,61
5	Letter Sets	0,52	0,17	0,22	0,53
6	Inference	0,62	-0,10	0,29	0,53
7	Cube Comparison	0,35	0,38	0,19	0,53
8	H.L. Reading Comp.	0,60	-0,09	0,11	0,36
9	Figure Class.	0,78	-0,02	-0,15	0,53
10	Blox	-0,01	0,76	0,11	0,60
11	Reasoning Ability	0,72	-0,02	-0,15	0,49
12	Pattern Comp.	0,58	0,28	-0,00	0,60
13	Figure Series	0,58	0,14	-0,06	0,43
14	Adv. Vocabulary	0,00	-0,05	0,76	0,58

TABLE 18

AFRIKAANS SAMPLE N = 72

FACTOR CORRELATION MATRIX

	I	II	III
I	1,00	-	-
II	0,59	1,00	-
III	0,33	0,11	1,00

6.2. Interpretation of Factors

It is interesting to note, that despite the differences between the intercorrelation matrices obtained for the English- and Afrikaans-speaking samples, the factor structures derived for these two groups are fairly similar. Indeed, they lead to surprisingly compatible factorial interpretations with only minor differences. The differences that are found to exist, it will be argued later, can be ascribed to differences in the problem-solving styles of the two samples, rather than to differences in their actual ability patterns.

But first let us examine the factors themselves, and attempt to give them psychological meaning. It will be seen by an inspection of Tables 13, 14, 15 and 17, that in the case of both language groups, the oblique, direct quartimin rotations give clearer simple structures. Interpretations of the factors will therefore be based on the oblique solutions. Close examination of the factor matrices will show, however, that the same interpretations are applicable to the uncorrelated factor solutions. The shifts in factor loadings caused by allowing the reference axes to become correlated, are in most cases small.

Considering the sample sizes used, a cut-off of 0,30 will be taken as a minimum loading for factor interpretation. Factor loadings falling just below this criterion will be considered only where these make good sense on psychological grounds. All lesser factor loadings will be treated as zero entries in the columns of the factor matrices.

Regarding Table 15, which gives the factor matrix obtained by direct quartimin rotation of the English sample factor matrix, the interpretable factor loadings in the first column are the following:

Figure Classification	0,68
Blox	0,68
Cube Comparison	0,64
Card Rotation	0,62
Deductive Reasoning	0,52
Letter Sets	0,41

Six tests have non-vanishing entries in this column. The item content of the first four listed tests obviously requires the manipulation of objects in space. Figure Classification was included in the battery as a reference for an induction factor, but the pictorial presentation of this test clearly calls for an imaginal manipulation of the stimulus materials. The interpretation of this factor can therefore be given with reasonable confidence as space. The appreciable loading of the D.R.T. on this factor is somewhat unexpected. A possible explanation could be that many of the subjects resorted to the use of Venn diagrams to aid in the solution of the syllogisms. This would, in effect, lead to a pictorial representation of the problems, which could then be solved spatially by an imaginal manipulation of the sets represented in the Venn diagrams. Evidence that this strategy was employed by many of the subjects was obtained by an inspection of the test booklets. Many of these were covered in little Venn diagram representations of the problems.

It will be seen shortly, that no evidence of a separate deduction factor was produced in the study. Since none of the selected marker tests for deduction appears to have been appropriate to the measurement of this construct, the D.R.T. had a high proportion of specific variance that could probably be attributed to the deduction component of the test. The remainder of its reliable variance was possibly due to a general reasoning component, and for this reason the test had moderate loadings on the other factors requiring reasoning. It should not, however, be concluded that the D.R.T. does not measure deduction.

It can be seen in Tables 15 and 17 that the communality of this test was only 0,44 in both language samples. If the test has a reliability of 0,90 it follows that 0,46 of its reliable variance is specific. This large specificity may well represent the deductive component of the test. But in view of the fact that none of the postulated deduction reference tests fulfilled their purpose, this variance could not be accounted for by a deduction factor. It should be borne in mind that the common and specific components of a test's reliable variance are determined by the relatedness of the instrument to other tests in the battery. They are not fixed properties of the particular test concerned. Had there been other good measures of deduction in the battery therefore, the communality of the D.R.T. would probably have been greater. The test could then have merged with the other deduction variables in a single factor. It is nevertheless unfortunate that no deduction factor could be produced in the present study.

The relatively small loading of Letter Sets on the space factor is not difficult to account for. It is possible to answer many of the easier items in this test by resorting to a spatial scanning of the material. The amount of inductive reasoning required in these items is then reduced to a minimum.

The second column in Table 15 has only three non-vanishing entries.

High Level Vocabulary	0,88
Advanced Vocabulary	0,88
High Level Reading Comprehension	0,31

The interpretation of this factor as verbal meaning should provoke little dispute. The three postulated reference markers for this factor have emerged very clearly. It is interesting to note that Reading Comprehension made a relatively minor contribution to the interpretation of the factor. This is to be expected as

Reading Comprehension tests have often been found to contain a component of general reasoning. The multidimensionality of the Reading Comprehension test can clearly be seen in the factor matrix in Table 15.

The third column in Table 15 has four non-vanishing entries.

Pattern Completion	0,75
High Level Reading Comprehension	0,54
Figure Series	0,48
Locations	0,48

The interpretation of this factor is less straightforward than in the case of the previous two. Pattern Completion is a matrix test, which presumably requires a combination of induction and general reasoning. Reading Comprehension, as discussed above, almost certainly requires general reasoning. Figure Series was included in the test battery as a postulated reference marker for a factor of deduction. This choice however, was made in the absence of any factorial evidence of the test's psychological identity. The inclusion of this test as a measure of deduction was, in fact, made on the basis of its structural similarity to the Form Series Test of Blakey[49)] (1941). It is very likely, however, that the test also requires a good deal of general reasoning as series tests have frequently been shown to load on this dimension. Locations requires the discovery of rules from an embedding context. This test can therefore be regarded as a measure of induction. It is also possible that a degree of general reasoning is required in the test. A tentative interpretation of this factor as some form of general reasoning is therefore offered. It should be recognised that a good deal of induction variance is accounted for by the factor too, and an alternative interpretation of the factor as induction should not be discounted. It can now be appreciated that there is a complete lack of evidence for a separate dimension of deduction.

Even when four factors were extracted no evidence for a factor of deduction was found. It should therefore be accepted that the tests selected to measure this elusive dimension in the present study, were not suitable.

To what extent is the factor structure obtained for the Afrikaans-speaking sample compatible with that described above?

Column 1 in Table 17 has no less than ten non-vanishing entries. These are listed below according to the magnitude of their factor loadings.

Figure Classification	0,78
Reasoning Ability	0,72
Deductive Reasoning	0,66
Inference	0,62
Location	0,61
High Level Reading Comprehension	0,60
Pattern Completion	0,58
Figure Series	0,58
Letter Sets	0,52
Cube Comparison	0,35

In seeking an interpretation for this factor it is almost simpler to begin with a consideration of those tests on which it does not have substantial loadings. The essentially zero entries of the two vocabulary tests, as well as the zero entries of the space tests, Card Rotation and Blox, confirm that this factor is not sufficiently broad to be denoted "g" or general intellectual efficiency. Among the variables having highest loadings on the factor are a mixture of hypothesized deduction and induction measures. Reading Comprehension also has a considerable loading on the factor. Cube Comparison, which, as mentioned earlier, had a bi-model frequency distribution in the Afrikaans sample, has shared its variance between this factor and the next one, which is a clear space factor. The multi-dimensionality of this test suggests that it can be responded to in two distinct stylistic fashions. It can either be approached perceptually, by way of imaginal manipulation, or it can be approached as an analytical reasoning task.

It seems that this very broad factor has its counterpart in the narrower general reasoning factor that was found in the English sample. An alternative label for the factor, as in the case of its English counterpart, could be analytical reasoning. The tests all require a careful analysis of the properties contained in the items before an answer can be produced by reasoning. This label suggests a stylistic contrast to traditional space tests, which are responded to on a more global, perceptual level. The question of different response styles, or problem-solving styles, will be discussed in more detail in a later section.

The second column in Table 31 has three interpretable entries.

Blox	0,76
Card Rotation	0,67
Cube Comparison	0,38

This is an unambiguous space factor. It has its counterpart in the English sample in the factor represented by the first column in Table 15. It is interesting to note that Figure Classification does not load on this factor in the Afrikaans sample whereas this test had the highest loading on the English space factor. A careful examination of the actual test protocols provided a valuable clue to explaining this otherwise curious result.

Figure Classification, it will be remembered, is a test consisting of 224 items. This is a formidable test length for any subject to be confronted with in an allotted time period of 18 minutes. Indeed, despite the fact that the time limit allowed in the administration of the test was two minutes longer than that prescribed in the E.T.S. test manual, most subjects were still unable to complete many of the items. A tally of the actual number of items completed, revealed that on the average, subjects attempted only 70% of the items in the test. But there were two distinct trends discernable in the way the test was approached. In the English-speaking sample, most subjects appear to have worked through the test quickly, omitting items all along. In fact, it seems they proceeded to work through the test attempting all the easier items, which could be solved quickly in a perceptual way. All the more difficult items, that required subjects to pause and reason in an analytical fashion, were omitted.

The Afrikaans-speaking subjects on the other hand, appear to have approached the test with a somewhat different strategy. The general trend observed in the test protocols of these subjects, suggests that they began at the first item and proceeded to work systematically through the test, attempting every item without regard to its difficulty. This resulted in their completing fewer items than their English counterparts, since they had to spend time lingering over difficult items. It also meant that their test protocols were characterized by large numbers of unattempted items at the end of the test, rather than having randomly interspersed omissions throughout the test. It seems likely that the lower mean of the Afrikaans sample on this test is due to the different stylistic approaches of the two groups. It should perhaps be mentioned in this respect, that the items in the test do not appear to be arranged in ascending order of difficulty.

From the foregoing, it is concluded that the general trend amongst the Afrikaans-speaking subjects was to approach Figure Classification as a test of analytical reasoning ability. This would account for the loading of the test on the analytical reasoning factor in this sample. The trend amongst the English-speaking subjects, on the other hand, appears to have been to treat the items in a global, perceptual way. This response style would explain the high loading of the test on the space factor in the

English sample. The apparent stylistic difference between the approaches of the two samples in doing the test could be responsible for the observed difference between the means of the two populations on the test. The English subjects of the average, were able to attempt more items in the test because they tended not to waste time on difficult items. Hence they were able to obtain higher scores. This is reflected in the higher mean score of this group.

The hypothesized difference in the cognitive styles of the two language groups, provides a possible explanation for all the major discrepancies between factor loadings observed in the two groups. It will be seen that those tests that have substantially different factor loadings in the two samples, viz. those tests of which the factor loadings are not invariant across the two populations, are pictorially or non-verbally presented tests which require a combination of perceptual-spatial skills and analytical-reasoning skills. In the English sample these tests tended to load relatively heavily on the space dimension, whereas in the Afrikaans sample they tended to have substantial projections on the general/analytical reasoning dimension. The third column in Table 17 has only two non-vanishing entries. A third entry falls just short of the criterion and may be considered in the interpretation of this factor. The three tests marking this dimension are presented with their respective factor loadings below:

High Level Vocabulary	0,76
Advanced Vocabulary	0,76
Inference	0,29

This factor is essentially a vocabulary doublet. The lesser entry of Inference suggests that the factor could be the broader verbal meaning dimension, since Inference contains verbally stated problems that require an understanding of the meaning of the material before correct inferences can be drawn. Its counterpart in the English sample is to be found in the second column of the oblique factor matrix in Table 15. Surprisingly, Reading Comprehension has no loading on this factor in the Afrikaans sample. It seems that these subjects treated Reading Comprehension purely as a test of general reasoning.

In summary of the above interpretations of the factors the following observations are relevant.

The expectation that a clearly defined factor of deduction would emerge was not borne out in the analyses. It appears that the selected reference markers for this factor were not appropriate to the identification of this construct. This is not altogether surprising. The basic assumption underlying the attempt to develop a test of deductive reasoning (Verster,[50] 1972) proposes that it is tenable to postulate that a psychological dimension in the form of deductive reasoning ability can be measured. Yet to date no attempts at measuring this ability have been successful. This is the situation with which the author (Verster,[51] 1972) was confronted at the outset of the investigation. A major explicit aim of the study was to attempt to rectify the situation by developing a useful, reliable and valid test of deductive reasoning. An operational model of the dimension was conceptualized, and a test with parameters based directly on the model was constructed. An analysis of the item content of the test shows that on operational, as well as on introspective grounds, the test requires reasoning of a nature commensurate with the definition of deduction. It seems probable therefore, that the high specific variance obtained for this test in the context of the present battery, reflects variance attributable to deductive reasoning ability. Since none of the other tests in the battery was capable of eliciting variance that could be ascribed to this ability, the deductive variance of the D.R.T. could not be accounted for by a deduction dimension. It consequently remained specific to the test. It is nevertheless of interest to note that the test appears to be multidimensional, sharing the remainder of its reliable variance between factors of spatial reasoning and general/analytical reasoning. The respective proportions of variance ascribed to these factors appear to be a function of the particular style of responding to the test, favoured by the sample tested.

Clearly determined, unambiguous factors of verbal meaning and spatial reasoning were found in both populations. A third factor which has tentatively been labelled general/analytical reasoning also appears to be common to both language groups. This factor was more strongly overdetermined in the Afrikaans sample. This is in contrast to the spatial reasoning factor which had a stronger overdetermination in the English sample. The tendency for certain pictorially or non-verbally presented reasoning tests to shift their factor loadings between these two factors is regarded as evidence of a difference in cognitive and problem-solving style between the two cultural groups. It is too early, at this stage, to say what factors are responsible for the development of different response styles in the two populations. Further research into this extremely interesting problem is strongly recommended.

6.3. Inter-Group Factor Analysis

In order to contrast the factor structures of the two language groups, an inter-group factor analysis procedure was employed. The method that was used in this analysis is an adaptation of Meredith's inter-group procedure (Meredith,[52] 1964 (b)). A description by Browne[53] (1969) of the mathematical formulation of the model and the rationale of the technique is given in Grant[54] (1969). This procedure was used in preference to the method of congruent factors of Cliff[55] (1966). The former technique offers greater flexibility in that it allows for within group differences in factor variances as well as factor covariances.

The assumption on which the inter-group procedure is based is that a single factor matrix common to all groups can be found. The observed differences between the factor matrices of the separate groups are accounted for in terms of differences in factor variances and covariances. The procedure can only be used with justification if all the variables from which scores are obtained are common to all groups. It was consequently necessary to exclude the hypothesized verbal meaning measures from the analysis since different tests were used to measure this dimension in the two language samples. The exclusion of these variables should not, however, have an adverse effect on the nature of the final structure that emerges. It will be recalled that in the separate factor solutions obtained for each language group (Tables 15 and 17), fairly neatly overdetermined verbal meaning factors emerged. In both solutions this dimension was more or less uncorrelated with the other factors. It can therefore be assumed with reasonable confidence that the dimensions of verbal meaning obtained in the factor structures of the two samples are congruent. The aggregate factor structure of the total population should, as a result, contain a common verbal meaning dimension. The exclusion of this dimension from the analyses should therfore not affect the interpretation of the aggregate factor structure to any great extent.

The variables that were excluded are the High Level Battery Reading Comprehension and Vocabulary tests for English and Afrikaans and the Advanced Vocabulary test with its counterpart, the Gevorderde Woordeskattoets. The remaining eleven tests in the battery are directly comparable in the two languages. Three rows and three columns of the sample intercorrelation matrices (Tables 9 and 10) corresponding to the excluded variables, were deleted. The reduced sample intercorrelation matrices were used in the inter-group factor analysis.

An iterative principle factor analysis procedure (Thomson[56] 1956) was used to obtain the separate factor matrices from the reduced intercorrelation matrices. The number of factors to be iterated on was specified by Kaiser's (1960, p. 146) criterion. This decision rule called for the extraction of two factors in each group. Squared multiple correlations were used for the initial communality estimates. Iteration was continued until the communalities converged within a tolerance of 0,0005. The unrotated principle factor matrices together with the varimax rotations, direct quartimin rotations and the communalities are given in Tables 19 and 20. The factor correlation matrices derived from the direct quartimin rotations are given in Tables 21 and 22.

The separate orthogonal factor matrices of the two groups served as input to the inter-group procedure. In terms of the procedure, the rotated orthogonal factor matrices of the separate groups were rescaled by means of the rescaled standard deviations (Cf. Tables 5 and 6) in order to obtain comparability of factor variances. An average factor matrix common to both groups was derived from the separate rescaled orthogonal factor matrices of each group. The method used to obtain the average matrix is that described by Meredith[57] (1964 (b)). The common factor matrix was rotated to an oblique solution by means of the quartimin procedure.

The obliquely rotated average factor matrix common to both groups appears in Table 23. The factor correlation matrix and the average factor covariance matrix appear in Tables 24 and 25 respectively. The average factor matrix served as a target (model) in rotating the separate factor matrices of the two groups to relative equivalence. The factor variances of the two groups appear in Table 26. The factors were correlated to a very similar degree in the two groups. In the English group the value was 0,65 and in the Afrikaans group it was 0,61. The factor pattern matrices of the two groups appear in Table 27.

TABLE 19

UNROTATED FACTOR MATRIX, VARIMAX ROTATION, DIRECT QUARTIMIN ROTATION AND COMMUNALITIES : ENGLISH SAMPLE N = 88

	Test	Unrotated Matrix		Varimax Rotation		Direct Quartimin Rotation		h^2
		I	II	I	II	I	II	
1	Locations	0,50	0,24	0,24	0,49	0,12	0,48	0,57
2	D.R.T.	0,62	0,06	0,44	0,44	0,34	0,35	0,44
3	Card Rotation	0,61	-0,27	0,64	0,17	0,68	-0,03	0,42
4	Letter Sets	0,62	0,21	0,35	0,55	0,23	-0,50	0,52
5	Inference	0,30	0,30	0,04	0,42	-0,08	0,46	0,41
6	Cube Comparison	0,49	-0,30	0,57	0,07	0,62	-0,11	0,50
7	Figure Class.	0,67	-0,17	0,63	0,28	0,63	0,11	0,50
8	Blox	0,64	-0,23	0,65	0,22	0,67	0,03	0,82
9	Reasoning Ability	0,45	0,36	0,12	0,56	-0,03	0,59	0,51
10	Pattern Comp.	0,64	0,12	0,42	0,50	0,33	0,42	0,59
11	Figure Series	0,49	-0,09	0,44	0,24	0,43	0,12	0,38

TABLE 21

FACTOR CORRELATION MATRIX (DIRECT QUARTIMIN ROTATION) : ENGLISH SAMPLE N = 88

	I	II
I	1,00	-
II	0,52	1,00

TABLE 20

UNROTATED FACTOR MATRIX, VARIMAX ROTATION, DIRECT QUARTIMIN ROTATION AND COMMUNALITIES : AFRIKAANS SAMPLE N = 72

	Test	Unrotated Matrix		Varimax Rotation		Direct Quartimin Rotation		h^2
		I	II	I	II	I	II	
1	Locations	0,75	0,04	0,64	0,39	0,65	0,15	0,58
2	D.R.T.	0,64	0,17	0,62	0,23	0,67	-0,02	0,44
3	Card Rotation	0,58	-0,30	0,30	0,58	0,20	0,52	0,42
4	Letter Sets	0,71	0,11	0,65	0,31	0,68	0,06	0,52
5	Inference	0,61	0,21	0,62	0,17	0,69	-0,09	0,41
6	Cube Comparison	0,71	-0,08	0,54	0,47	0,51	0,29	0,50
7	Figure Class.	0,68	0,21	0,67	0,22	0,74	-0,06	0,50
8	Blox	0,65	-0,63	0,18	0,89	0,04	0,93	0,82
9	Reasoning Ability	0,67	0,24	0,69	0,19	0,76	-0,10	0,51
10	Pattern Comp.	0,77	-0,03	0,61	0,46	0,60	0,25	0,59
11	Figure Series	0,62	0,03	0,53	0,32	0,54	0,13	0,38

TABLE 22

FACTOR CORRELATION MATRIX (DIRECT QUARTIMIN ROTATION) AFRIKAANS SAMPLE N = 72

	I	II
I	1,00	-
II	0,57	1,00

TABLE 23

AVERAGE FACTOR MATRIX (DIRECT QUARTIMIN ROTATION)

	Test	I	II
1	Locations	0,57	0,11
2	D.R.T.	0,52	0,16
3	Card Rotation	0,06	0,62
4	Letter Sets	0,61	0,11
5	Inference	0,62	-0,11
6	Cube Comparison	0,23	0,46
7	Figure Class.	0,42	0,30
8	Blox	-0,07	0,84
9	Reasoning Ability	0,71	-0,10
10	Pattern Comp.	0,51	0,27
11	Figure Series	0,32	0,28

TABLE 24

FACTOR CORRELATION MATRIX (DIRECT QUARTIMIN ROTATION)

	I	II
I	1,00	-
II	0,62	1,00

TABLE 25

FACTOR COVARIANCE MATRIX

	I	II
I	1,00	-
II	0,62	1,00

TABLE 26

FACTOR VARIANCES

	VARIANCES	
	FACTOR I	FACTOR II
ENGLISH	0,65	0,86
AFRIKAANS	1,42	1,17

TABLE 27

ROTATED FACTOR PATTERN MATRICES

		ENGLISH		AFRIKAANS	
	Test	Analytical Reasoning	Space	Analytical Reasoning	Space
1	Locations	0,69	-0,03	0,49	0,23
2	D.R.T.	0,56	0,26	0,51	0,05
3	Card Rotation	0,05	0,66	0,08	0,57
4	Letter Sets	0,67	0,05	0,58	0,16
5	Inference	0,54	-0,19	0,65	-0,01
6	Cube Comparison	-0,07	0,60	0,38	0,38
7	Figure Class.	0,25	0,63	0,54	0,02
8	Blox	0,12	0,58	0,19	1,06
9	Reasoning Ability	0,77	-0,20	0,66	-0,02
10	Pattern Comp.	0,62	0,19	0,44	0,33
11	Figure Series	0,24	0,41	0,37	0,18

In order to determine the goodness of fit of the model the square root of the average squared residual was calculated for each group separately. For the English group this statistic yielded a value of 0,092. For the Afrikaans group the value of this statistic was 0,090. The square root of the average of all the squared residuals was 0,091. From these statistics it would seem that the model afforded an acceptable fit to the data.

6.4. Discussion

In the following discussion attention will firstly be given to the effects of excluding the variables measuring the verbal meaning dimension. A comparison will be made between the three factor and two factor solutions for the separate groups. The two factor solutions obtained in each group will also be compared with one another. An attempt will be made to account for the differences between the factor structures of the two groups. The discussion will then proceed to a consideration of the average factor matrix derived for the total CSIR population.

An examination of the direct quartimin rotations in Tables 19 and 20 reveals that reasonably compatible structures were obtained for the English and Afrikaans groups. Factor one in the English sample is a clearly determined space factor. The second factor has its highest loadings on the tests selected to measure both deduction and induction. The factor can perhaps be labelled deductive-inductive reasoning. A better description for this factor though, would be general reasoning or analytical reasoning. The first oblique factor in the Afrikaans group (Cf. Table 20) also has its highest projections on a combination of hypothesized deduction and induction measures. It is the same kind of general/analytical reasoning factor as the second dimension in the English group. The second factor in the Afrikaans group is a neatly determined space factor corresponding to the first factor in the English group. The relatively low loading of Cube

Comparison on this dimension can probably be attributed to the bi-modal nature of the frequency distribution of scores on this test in the Afrikaans sample. There were clearly two styles of responding to this test in the Afrikaans group, as evidenced in the positive loadings of this test on both factors.

It is interesting to note that, whereas the space factor accounted for the greatest proportion of the variance in the English sample intercorrelation matrix, the general/analytical reasoning factor accounted for the greatest amount of variance in the Afrikaans data. This is an important observation, for it appears to throw light on the nature of the preferred style of responding to ability tests characterizing the two populations. This point will be discussed in greater detail at a later stage.

A comparison between the oblique factor solutions in Tables 19 and 15 and Tables 20 and 17 respectively, reveals some interesting findings. In the English sample two-factor and three-factor solutions (Cf. Tables 19 and 15) there is considerable overlap. The Factors I and III in Table 15 are, with only minor differences in factor loadings, very similar to Factors I and II respectively in Table 19. This implies that the exclusion of the three verbal meaning tests had little effect on the nature of the factor structure in this sample. Their exclusion has merely resulted in the loss of the verbal meaning factor, leaving the nature of the remaining two factors unchanged. It is of interest to observe, however, that the magnitude of the intercorrelations between the two remaining factors is increased when the verbal meaning dimension is not included in the analysis. The correlation between space and general/analytical reasoning in the three-factor solution is 0,44 (Cf. Table 16). These factors correlate to the extent of 0,52 in the two-factor solution (Cf. Table 21). Similar results can be observed in a comparison of the two-factor and three-factor solutions for the Afrikaans data. Factors I and II in Table 17 are very similar to Factors I and II in Table 20. Again, the exclusion of the three verbal meaning tests has merely resulted in the absence of a verbal meaning dimension in the two-factor solution, without further affecting the structure of the remaining factors. The intercorrelation between the space and general/analytical reasoning factors has remained relatively constant in this case, despite the exclusion of the verbal meaning dimension.

From the results it can be concluded that the intellectual structure in the case of both populations is characterized by highly co-operative factors of spatial reasoning and general/analytical reasoning, together with a relatively independent dimension of verbal meaning. It should be remembered in this regard, that other factors are also likely to be found in the ability structures of the two populations, but these were not produced in the present study due to a limited sample of ability tests in the battery. It is notable that even with the more refined principal factor analysis technique used to obtain the two-factor solutions, the variance accounted for by the postulated measures of deduction and induction could not be separated. A number of alternative explanations could account for this finding. It may simply be that the tests chosen to measure these dimensions were not appropriate to the task. It should be remembered, that with the exception of the D.R.T., these tests were all constructed for use with high school and college student populations. The formats of these tests may not be suitable for eliciting separate sources of variance in a more sophisticated population. It is also possible that the tests were too highly speeded for the general CSIR population. If this is the case, the merging of deduction and induction into a common factor could be ascribed to a high percentage of common speed variance in the tests.

Attention will now be focussed on an examination of the differences between the two-factor solutions obtained for the separate groups. This will be accomplished by means of a comparison of the direct quartimin rotations in Tables 19 and 20. In terms of the program that was used to perform the analyses, the factor that accounts for the greatest proportion of the variance in the intercorrelation matrix is printed first. Each subsequent factor that is printed accounts for progressively less variance. Inspection of Tables 19 and 20 reveals that the space factor accounted for the greatest proportion of the variance in the English test battery while the general/analytical reasoning factor accounted for most of the variance in the Afrikaans test battery. As a result the space factor was better determined in the English data and the general/analytical reasoning dimension was more strongly overdetermined in the Afrikaans data. The hypothesis advanced in an earlier section, that the lack of factorial invariance of the test loadings can be attributed to differences in the problem-solving styles of the respective populations, appears to be supported by this finding. The major shifts in factor loadings across the two populations occur in respect of Figure Classification, Cube Comparison and Figure Series. These three tests are all presented in a pictorial medium and require a combination of perceptual/spatial skills and analytical reasoning skills. In the English factor matrix these tests all had their highest loading on the space dimension whereas in the Afrikaans matrix the highest loadings of these tests were on the general/analytical reasoning dimension. It would thus appear that the stimulus material in these tests permits the production of a solution to the items in terms of two different cognitive styles. The cognitive style, or problem-solving style adopted by the English-speaking subjects in doing these tests, appears to be of a global, perceptual, nature. The Afrikaans-speaking subjects on the other hand, tended to favour a style characterized by systematic, analytical reasoning. These broad stylistic preferences could explain the differential amounts of variance accounted for by the two factors in each group. It is also possible that the observed differences in the mean test scores of the two populations are due to differences in the problem solving styles of the two groups.

The average factor matrix common to both language groups will now be examined. The average factor matrix, rotated to an oblique position by the direct quartimin procedure, appears in Table 23. An inspection of the factor loadings in the first column of the matrix reveals that the highest loadings are on Reasoning Ability, Inference, Locations and Deductive Reasoning (D.R.T.). Other tests with significant loadings on this dimension are Pattern Completion, Figure Classification and Figure Series. These tests were all chosen as references for either deduction or induction. The tests with the highest significant projections on the second factor are Blox, Card Rotation and Cube Comparison. Figure Classification has a marginally significant loading on this factor. It is a uniquely determined, unambiguous space factor. It is interesting to note that Figure Classification had a much higher loading on the space factor for the English sample.

An unequivocal interpretation for the first factor is difficult to find. Since the tests defining the factor call for both deductive and inductive reasoning processes, it is likely that some form of rational-analytical thinking ability is being measured. The label that has been used to describe this factor in discussing the factor matrices of the separate groups, namely general/analytical reasoning, is perhaps too broad. Traditionally, factors of general reasoning are characterized by tests of arithmetical and mathematical reasoning and figure analogies. These tests require reasoning of a broader, more general nature than the deductive and inductive tests in the battery. Since these tests all call for a rational analysis of the material in the items before inferences can be made, whether deductive or inductive, a more descriptive name for the factor might be analytical reasoning. It is not known to what extent the reasoning required in these tests correspondes to the type of analytical reasoning required in embedded figure tests. It seems likely that tests of the latter kind rely more heavily on skills in perceptual analysis, while the tests of deduction and induction rely on skills in rational analysis. Another plausible description for this factor might be inferential reasoning, since the drawing of inferences is common to both inductive and deductive reasoning processes. The term analytical reasoning however, is preferred as it is considered to be more aptly descriptive. In conclusion, it may be observed that the average factor structure representing the pattern of abilities of the total CSIR population, bears a close resemblance to the factor matrices of the separate language populations. The analysis was not able to separate deductive and inductive variance, although a separate dimension of spatial reasoning emerged very clearly. It should be noted that the two factors that emerged were fairly highly correlated (Cf. Table 24). This finding could imply that a second-order analysis would produce a single general factor. Nonetheless, the first-order factors that emerged are regarded as being sufficiently neatly determined to warrant serious consideration as primary factor abilities.

7. MULTIDIMENSIONAL SCALING

7.1. Description of the Procedure

A multidimensional scaling procedure in the form of a Principal Co-ordinates analysis was undertaken with a view to examining the pattern and location of the test vectors of the tests in the battery in a multidimensional space. The procedure used in the analysis was derived from the method of Gower[58] (1967). Separate analyses were performed on the English sample covariance matrix (Cf. Table 11) and on the Afrikaans sample covariance matrix (Cf. Table 12). All fourteen variables in the battery are included in the analyses.

According to the Principal Co-ordinates procedure, the covariance matrix is rescaled by means of the number of items in the tests. This is necessary in order to weight the covariances in terms of the number of items in each test, since a test with many items is likely to have a greater standard deviation than a test with few items. The rescaled matrices are used as input to the program. The matrices of principal components of the rescaled input matrices are then obtained.

The first three columns of the obtained principal component matrices are shown in Tables 28 and 29 respectively. The first principal component to be extracted in the analysis accounts for the greatest proportion of the variance. Each successive principal component accounts for progressively less variance. An examination of the latent roots indicates the declining proportion of variance accounted for by successive principal components. The latent roots corresponding to the two principal component matrices appear in Tables 30 and 31. Only the first three columns of the matrices are considered worthy of interpretation.

Since a principal component matrix is difficult to interpret in numerical form, the principal co-ordinates are presented in graphical form. In the case of each matrix separate plots of the first two columns, and of the first three columns respectively, were made.

The two and three dimension plots from the English sample matrix appear in Figures 1 and 2 respectively. The corresponding plots for the Afrikaans sample matrix appear in Figures 3 and 4. The points in the space representing the test vectors are indicated by the identity numbers of the tests.

7.2. Discussion

The results of the analysis performed on the English sample rescaled covariance matrix will be discussed first. Regarding the plot of the first two columns of the principal component

TABLE 28

ENGLISH SAMPLE

PRINCIPAL COMPONENTS MATRIX

Test		Principal Components		
		I	II	III
1	Locations	0,11	0,31	0,88
2	D.R.T.	-0,05	-0,47	-0,46
3	Card Rotation	-0,60	-0,49	0,03
4	H.L. Vocabulary	1,23	0,16	-0,29
5	Letter Sets	-0,13	-0,10	0,13
6	Inference	0,25	0,00	0,13
7	Cube Comparison	-0,55	-0,70	0,23
8	H.L. Reading Comp.	0,46	0,37	0,17
9	Figure Class.	-0,23	-0,42	-0,49
10	Blox	-0,20	-0,21	0,06
11	Reasoning Ability	0,17	-0,18	0,43
12	Pattern Comp.	-0,82	0,90	0,23
13	Figure Series	-0,72	0,78	-0,79
14	Adv. Vocabulary	1,09	0,07	-0,25

TABLE 29

AFRIKAANS SAMPLE N = 72

PRINCIPAL COMPONENTS MATRIX

Test		Principal Components		
		I	II	III
1	Locations	-0,37	-0,51	-0,29
2	D.R.T.	-0,15	-0,52	-0,62
3	Card Rotation	-0,46	-0,65	0,74
4	H.L. Vocabulary	0,86	0,02	0,14
5	Letter Sets	0,06	0,11	-0,03
6	Inference	0,30	0,08	-0,30
7	Cube Comparison	-0,13	-0,18	0,30
8	H.L. Reading Comp.	0,21	0,13	-0,15
9	Figure Class.	-0,14	-0,01	-0,29
10	Blox	-0,01	-0,10	0,47
11	Reasoning Ability	-0,22	0,07	-0,37
12	Pattern Comp.	-0,66	0,77	0,13
13	Figure Series	-0,32	0,71	0,10
14	Adv. Vocabulary	1,02	0,08	0,18

TABLE 30

ENGLISH SAMPLE N = 88

LATENT ROOTS
4,97
2,89
2,36
2,01
1,61
1,25
1,07
0,93
0,83
0,70
0,56
0,41
0,39
0,00

TABLE 31

AFRIKAANS SAMPLE N = 72

LATENT ROOTS
2,91
2,14
1,74
1,30
1,15
1,05
0,90
0,75
0,66
0,61
0,51
0,39
0,27
0,00

<u>FIGURE 1</u>

<u>ENGLISH SAMPLE.</u>

<u>FIRST TWO COLUMNS OF PRINCIPAL COMPONENTS MATRIX.</u>

1. Locations.
2. Deductive Reasoning.
3. Card Rotation.
4. High Level Vocabulary.
5. Letter Sets.
6. Inference.
7. Cube Comparison.
8. High Level Reading Comprehension.
9. Figure Classification.
10. Blox.
11. Reasoning Ability.
12. Pattern Completion.
13. Figure Series.
14. Advanced Vocabulary.

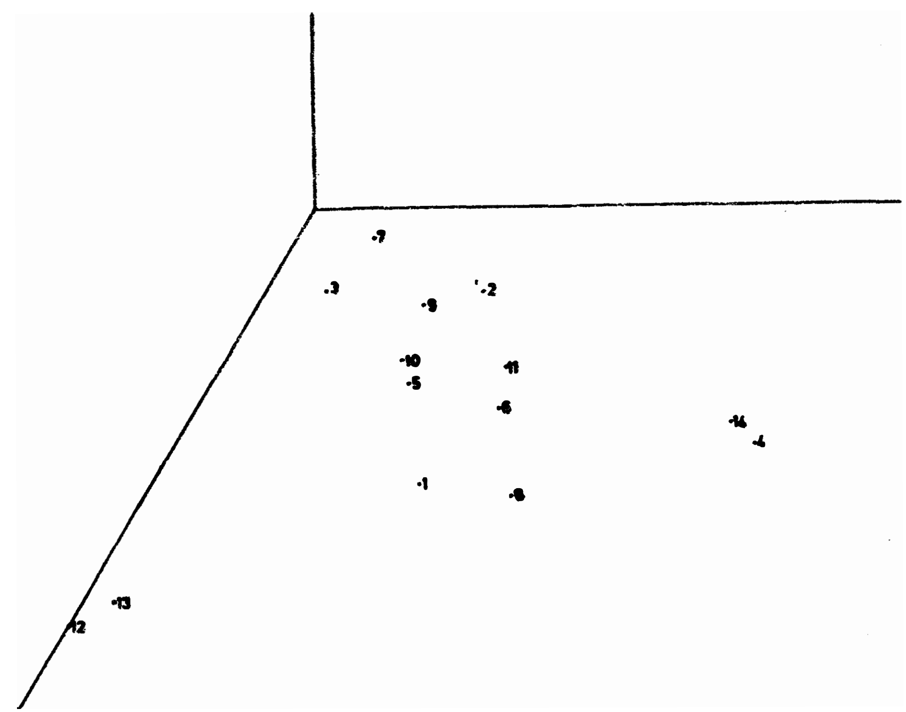


FIGURE 2

ENGLISH SAMPLE.

FIRST THREE COLUMNS OF PRINCIPAL COMPONENTS MATRIX.

1. Locations
2. Deductive Reasoning.
3. Card Rotation.
4. High Level Vocabulary.
5. Letter Sets.
6. Inference.
7. Cube Comparison.
8. High Level Reading Comprehension.
9. Figure Classification.
10. Blox.
11. Reasoning Ability.
12. Pattern Completion.
13. Figure Series.
14. Advanced Vocabulary.

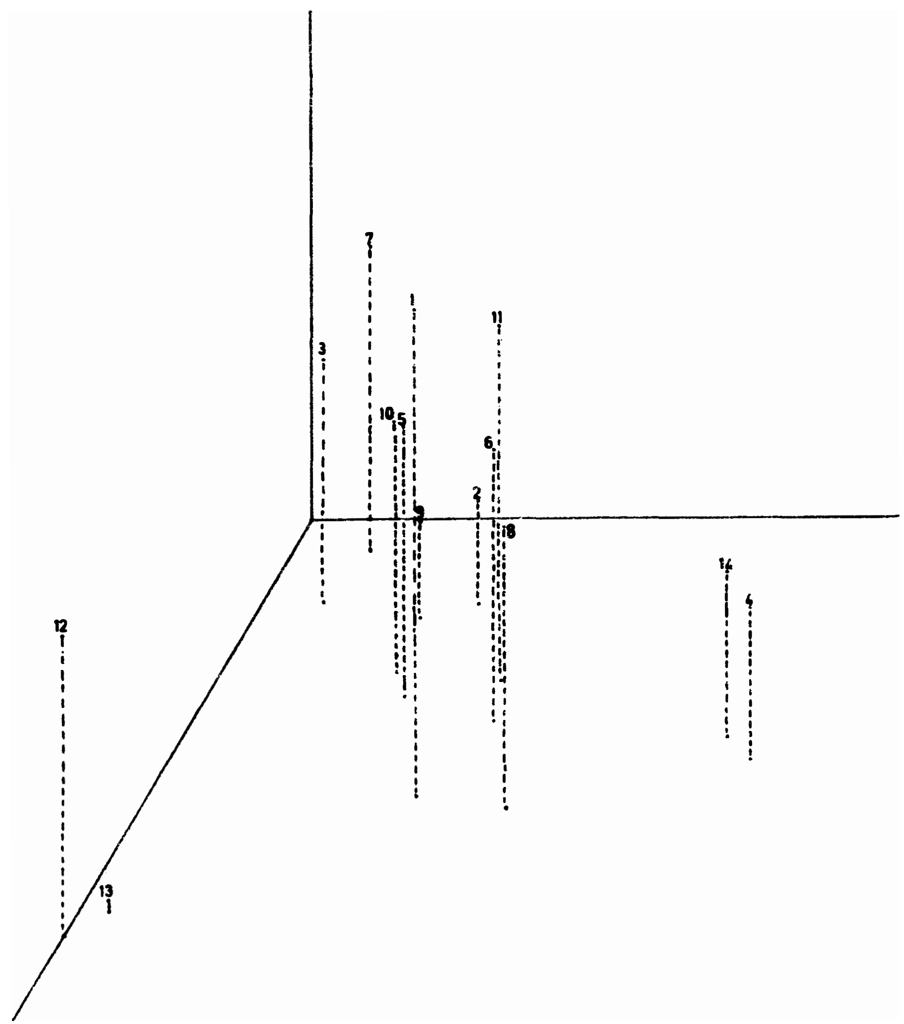


FIGURE 3

AFRIKAANS SAMPLE.

FIRST TWO COLUMNS OF PRINCIPAL COMPONENTS MATRIX.

1. Locations.
2. Deductive Reasoning.
3. Card Rotation.
4. High Level Vocabulary.
5. Letter Sets.
6. Inference.
7. Cube Comparison.
8. High Level Reading Comprehension.
9. Figure Classification.
10. Blox.
11. Reasoning Ability.
12. Pattern Completion.
13. Figure Series.
14. Advanced Vocabulary.

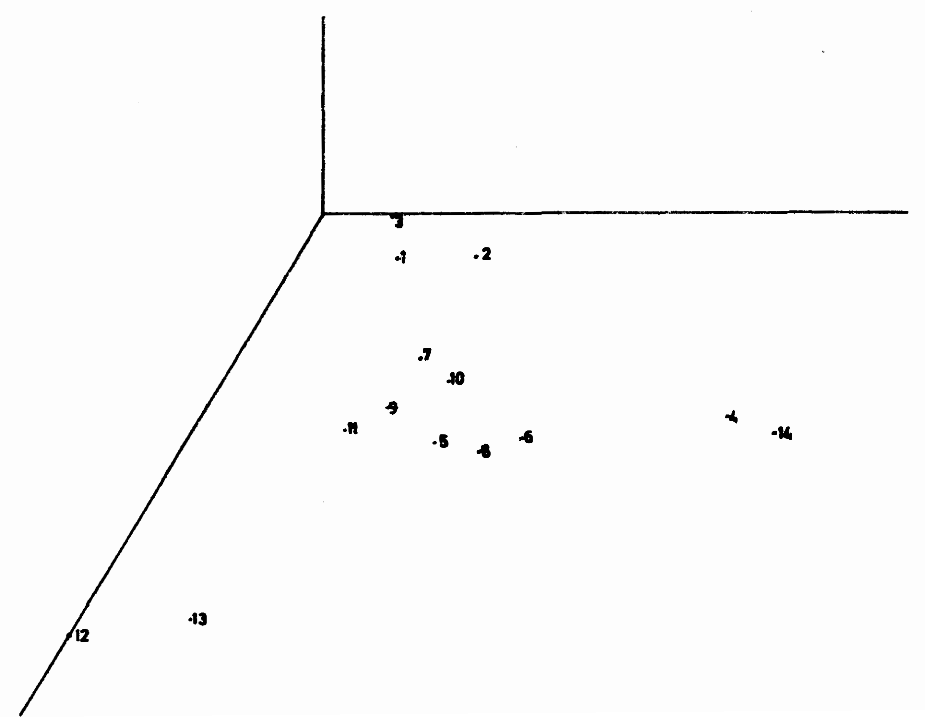
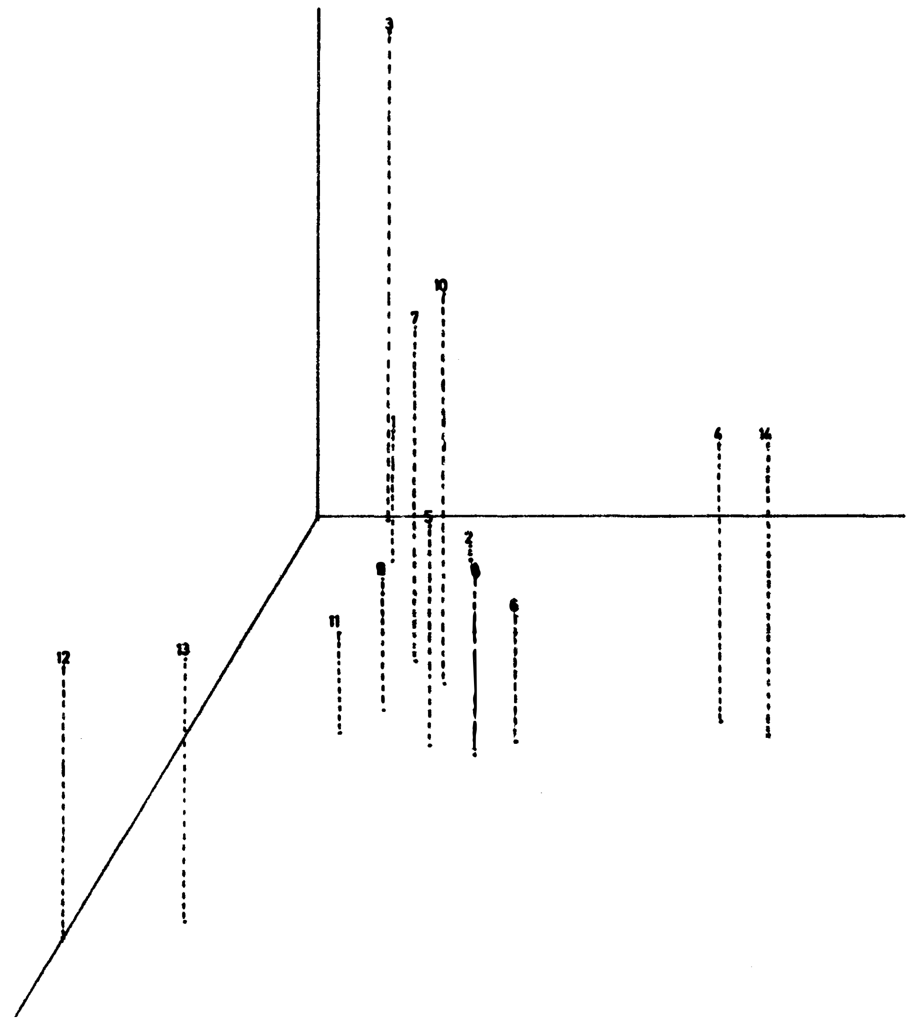


FIGURE 4

AFRIKAANS SAMPLE.

FIRST THREE COLUMNS OF PRINCIPAL COMPONENTS MATRIX.

1. Locations.
2. Deductive Reasoning.
3. Card Rotation.
4. High Level Vocabulary.
5. Letter Sets.
6. Inference.
7. Cube Comparison.
8. High Level Reading Comprehension.
9. Figure Classification.
10. Blox.
11. Reasoning Ability.
12. Pattern Completion.
13. Figure Series.
14. Advanced Vocabulary.



matrix in Figure 1, two distinct vectors of points can be discerned. All six points falling in the right half of the configuration represent tests administered by means of a verbal-medium. The remaining eight test vectors, represented by the points in the left half of the configuration, are all non-verbally presented tests. There is a distinct, non-overlapping dichotomy between the sets of points. The same picture can be discerned when the points of the test vectors are projected onto a three dimensional space. Unfortunately, the representation of a three dimensional space on two dimensional graph paper is not easy to attain. Nonetheless, it can still be observed that the points lying in the right half of the configuration represent verbal medium tests, while the points in the left half represent non-verbal or pictorial tests. It is interesting to note that the two points lying in the extreme right-hand corner of the configuration represent the two vocabulary tests. These tests were identified in the factor analyses as the purest measures of the verbal meaning dimension. The two points lying in the extreme left-hand corner of the configuration represent Pattern Completion and Figure Series. The distance between these two points and the points in the rest of the configuration, indicates that the two tests share relatively little in common with the rest of the battery. It seems likely that these tests rely more heavily on purely perceptual skills.

An examination of the results obtained from the analysis of the Afrikaans rescaled covariance matrix reveals a remarkably similar picture. An inspection of the two and three dimensional plots in Figures 3 and 4 respectively, reveals that once again the points of the test vectors are grouped according to the medium of presentation employed in the tests. The right half of the configuration constitutes the points representing verbally presented tests. The left half is characterized by the points of the non-verbal and pictorial test vectors. The dichotomy between the two sets of points is not quite as apparent as in the case of the English data.

One verbal-medium test, Reasoning Ability, falls in the left-hand plane of the configuration. In addition, the visual representation is rendered less clear due to the closeness of the points constituting the segments that fall towards the centre of the configuration.

The overall picture attained is nonetheless convincing. It is remarkably compatible with the picture represented by the configuration of points of the English sample test vectors. Once again the two vocabulary tests fall to the extreme right, while Pattern Completion and Figure Series lie in the extreme left-hand plane of the configuration.

It may be concluded in the light of these analyses, that performance on ability tests is directly affected by the medium in which the test items are presented. In the present study the tests were found to cluster in a multidimensional space according to the medium in which the items are communicated. One cluster of points represented tests presented by means of a verbal medium. The other cluster represented tests in which the medium for presenting items is either non-verbal or pictorial. These findings will be discussed in the following chapter in the light of Guttman's[59] (1969) facet model of ability tests. It should be noted in the present context, that analyses using Guttman's smallest space analysis technique were also performed (Schlesinger and Guttman,[60] 1969; Lingoes,[61] 1964). This is a non-metric scaling procedure in which the assumptions are less stringent than in the case of the principal co-ordinates analysis. Strikingly similar results were obtained. Indeed, the graphically represented pictures that emerged, were to all intents and purposes identical to those discussed above. Since the results obtained by means of the principal co-ordinates analysis are based on more stringent statistical assumptions, only these results are reported.

8. DISCUSSION

8.1. The Construct Validity of the Deductive Reasoning Test

A major objective of the study was to determine the construct validity of the Deductive Reasoning Test by means of factor analysis. It has been seen that the battery of selected tests failed to produce a factor of deduction. All the postulated reference tests for deduction merged in a single factor with tests of induction. The Deductive Reasoning Test was found to have a moderately high loading on this dimension both in the separate sample factor matrices and in the average factor matrix of the total population. From this it would appear that the D.R.T. shares some of its reliable variance with tests measuring both inductive and deductive processes. The interpretation that was offered for the induction-deduction dimension stressed that rational analytical reasoning was required in both types of tests. It appears then that this type of thinking is of importance in making deductive inferences. It should also be remembered that a considerable portion of the reliable variance in the D.R.T. was found to be specific to the test in the context of the experimental battery. This finding was taken to indicate that a significant proportion of the test's reliable variance may be attributed to deduction. This contention is founded in the strategy that was followed in constructing the instrument. It has been shown that the items in the test provide an operational definition of the conceptual model of deduction (Verster,[62] 1972). In order to arrive at the correct solution to any particular item in the test it is necessary to reason within the framework of deductive processes. Regarded in this light, it would be difficult to deny that the test contains a component of deductive reasoning. Although the study was not conclusive in its attempt to confirm the hypothesized factor structure of the test, it may still be maintained that on operational grounds the test provides a measure of the intended construct.

It is possible that the factor study was obstructed in its initial objective due to the poor discrimination power of some of the reference variables. Many of the crucial tests were not successful in measuring individual differences reliably. Their poor reliabilities are ascribed to a restricted spread of test scores. This in turn, may be attributed partly to the homogeneity of the pre-selected samples. The test lengths and small sample sizes are also likely to have had a significant effect in this respect. The D.R.T. in contrast, proved most successful in measuring individual differences in the highly sophisticated samples. It may be concluded that the test is appropriately tailored to the population for which it is intended. It should prove a valuable assessment tool in a selection battery for use with a highly homogeneous, intellectually sophisticated population.

8.2. The Validity of the Thurstonian Model in the CSIR Population

The second broad aim of the study was to demonstrate the generality of the Thurstonian model of intellect. This was to be achieved by seeking evidence for its appropriateness to the intellectual structure of the highly sophisticated CSIR population. From the results that were discussed in the last chapter, it is evident that there was at least partial support for Thurstone's conceptualization of the structure of intellect. Two of the three factors that emerged, could be identified with Thurstone's Primary Mental Abilities. These were space and verbal meaning. The third factor, that was tentatively interpreted as analytical reasoning, is of a much broader nature. It appears to cover general competency over a wide range of structured reasoning tasks. It is not known whether this factor is indicative of a real ability in the intellectual structure of the CSIR population, or whether it is merely an artifact produced by inadequate tests.

It should be remembered that it was stressed at the outset that the findings which emerge from the analyses could only be treated tentatively. It is possible that, had the reference tests been more appropriate to the measurement of individual differences in the experimental population, separate factors for deduction and induction would have emerged. Another criticism that may be justly levelled against the instruments in the study is that most of the reference tests were too speeded. This was found to be the case even though the author took the precaution of extending the prescribed time limits. The generally high reliabilities claimed for these tests by their authors probably reflect a large percentage of reliable speed variance. The presence of speed variance in ability tests usually results in spuriously inflated estimates of reliability. It would have been preferable to have had less speeded tests in the present study with higher reliabilities. It would appear from the findings, that future research in this domain would profit more if new and more appropriate tests are developed.

8.3. The Relation of the Findings to other Research

It may be contended that certain extremely interesting issues arose from the findings. The mean test score differences found between the English- and Afrikaans-speaking samples pose stimulating questions for future research. The experimental design in the present study does not permit conclusive inferences to be drawn in explanation of this result. But the differences that were found between the factor structures of the two populations do suggest that stylistic preferences in problem-solving behaviour might account for this observation. The finding that stylistic differences produced differences in the factor patterns of the two groups, is not entirely contrary to the predictions of the Thurstonian model. Thurstone[63] (1947) recognised that the stability of a pattern of intercorrelations among tests measuring higher mental processes is dependent on sampling fluctuations. He stated that "factor loadings cannot be expected to be invariant from one population to a different population". (Thurstone,[64] 1947, p. 360.)

Balinsky[65] (1941) demonstrated that large age differences will bring about differences in the factor structure of tests. He confirmed the classical example in which a test of addition is shown to be a reasoning exercise to fourth graders while it is a speed task to college students. Bloom and Broder[66] (1950, Ch. 3) and Lucas[67] (1953) have shown that large age differences are not the only considerations that affect the nature of the factor being measured by a test. Some tests measuring higher mental processes are solved in one way by some subjects and in a different way by other subjects. This implies that the tests may be measuring different abilities in the two groups of subjects. It follows then that the intercorrelations between the tests will be different for different groups of subjects. Since factor loadings are dependent on correlations, they cannot be expected to be invariant across different populations. A study was undertaken by French[68] (1965) in an attempt to determine the extent to which factor analysis is able to reflect qualitative differences in reactions to tests as well as to differences in the nature of the tests themselves. The study has a direct bearing on the present investigation, as many of the fifteen tests used by French have been included in the present study. Since there is also considerable overlap between the findings in the two studies, French's investigation will be discussed in some detail.

French administered a battery of fifteen tests selected to measure five important abilities, to a sample of 177 college students. The abilities he was interested in measuring were verbal comprehension, general reasoning, space, induction and visualization. One "pure" factor test was selected to represent each of the five factors. The remainder of the battery comprised ten tests selected to load these factors less purely, or to straddle two or more of them.

In order to determine what problem-solving styles the subjects used, they were asked to complete a questionnaire about their background and their approach to test problems. They were also interviewed while they solved sample items similar to the ones they had attempted in the tests. A procedure was devised whereby the interview and the questionnaire were quantified. The intercorrelations between the interview, the questionnaire and the test variables were then factor analysed. On the basis of the results, subjects were grouped into 17 pairs of subsamples representing different problem-solving styles, or different background characteristics that might be expected to affect the way in which a person solves problems. Separate factor analyses of the same fifteen tests were then performed in respect of all subsamples in the 17 pairs and of the whole group. In each of the 35 factor analyses five factors were extracted and rotated. Comparisons were made between the factor loadings and factor intercorrelations for the two subsamples in each pair.

The results showed clearly that many of the 17 subsample divisions could be loosely classified into a category called systematizing or analysing vs. scanning. French[69] (1965) considered that there might be a relationship between these broad classes of problem-solving style and the cognitive style of focussing vs. scanning that has been found to affect perception of size and susceptibility to illusions (Gardner et al.,[70] 1959). On the basis of the findings regarding the problem-solving style described as analytical, French[71] (1965) concludes that the concept of analytical attitude or field independence described by Witkin et al.[72),73)] (1954, 1962) relates to something quite different. It was found that many of the subsamples resorting to an analytical problem-solving style in certain tests, notably Concealed Figures and Cubes, obtained lower means than subsamples using a global-scanning style in these tests. Yet studies have shown that field independence, or analytic attitude correlates positively with general intelligence. It is interesting to observe the close correspondence between French's findings and those of the present study. In the present investigation it was also found that the sample using an analytical approach to Cube Comparison (Afrikaans sample) had a lower mean than the (English) sample in which the approach was of a global-scanning nature. French also found that in the analytic sub-sample the loading of Cubes on a space-visualization factor dropped from 0,52 to 0,07. He concludes that an analytic approach toward Cubes "destroys the capacity of that test to measure spatial ability". (French,[74] 1965, p. 22). This observation appears to be borne out in the present study.

In summarizing the results from his investigation, French comes to the conclusion that:

"tests, even simple "pure-factor" tests, do not measure the same things for all people. The kind of behaviour most readily observed in someone taking these tests was the use of some kind of reasoned or systematic approach as contrasted to less orderly scanning and visualizing, with reliance on common sense. This overall difference in problem-solving styles is the one emphasized by Bloom and Broder[75] (1950), (see especially Appendix), and may be related to Gardner's[76] (1959) Focussing vs. Scanning or to Witkin's[77] (1962) Analytic Attitude or Field Independence. For some of the tests, differences of this kind in test taking behaviour have no relation to the test's factorial content. For a few tests, the principal loading was strengthened. Most often, however, the use of a system in solving a test reduced the usual factor loading of that test. This happened more readily for spatial or visualization tests than for reasoning or verbal tests." (French,[78] 1965, pp. 26 - 27).

The appropriateness of French's conclusions to the findings in the present study, provides decisive support for the hypothesis that was offered earlier to account for the differences between the factor structures of certain tests in the two language groups.

8.4. Conclusion

From the foregoing discussion, it may be concluded that the validity of the Thurstonian model with respect to the intellectually sophisticated CSIR population, was partially confirmed. Of the four separate factors that were postulated, two emerged clearly and may be identified with the well established primary abilities of space and verbal meaning. The third broad factor that emerged, suggests that deductive and inductive processes are governed by a single ability in the form of analytical reasoning. The differentiation-integration theory of the growth of abilities, proposed by Lienert and Crott[79] (1964) may be able to account for this circumstance. These investigators propose that, during the growth of the individual to mental maturity, separate abilities develop differentially and each eventually becomes crystallized at a relatively invariant level. Further growth in the organism is brought about by the acquisition of experience and knowledge in specific domains and this may result in certain separately formed abilities becoming integrated in the form of single, broad competency factors. It is not difficult to imagine that, in the present investigation, the merging of deduction and induction tests in a single broad factor, reflects a similar integration of separately formed abilities. The heavy emphasis placed on both inductive and deductive reasoning processes in most scientific subjects at the universities, could perhaps be responsible for bringing about an integration of these separate abilities in a single rational-analytical reasoning factor.

The differences in the factor loadings of certain tests in the two groups formed on the basis of home language, may be reconciled in terms of differences in the problem-solving styles adopted by the respective samples. Subjects in the English-speaking sample tended to favour an approach characterized by a global-scanning style. The Afrikaans-speaking subjects on the other hand, showed a general preference for a systematic-analytic problem-solving style. It is likely that differences in the language-cultural back-grounds of the two populations are responsible for fostering differences in problem-solving style. An unpublished investigation conducted by Steyn in 1969 at the NIPR may be able to throw some light on this issue. Two large samples of young matriculated males were selected at random from a broad spectrum of socio-economic backgrounds in South Africa. One sample comprised 486 English-speaking subjects, the other comprised 489 Afrikaans-speaking subjects. On the basis of comprehensive biographical data, it was established that the samples were well matched on a number of relevant cultural and socio-economic parameters. Highly significant differences between the samples were found in respect of a personality dimension described as Over Regimentation. The results indicated that the Afrikaans-speaking subjects were more highly disciplined, systematic and regimented than their English counterparts. It is possible that these differences in personality and cognitive style, reflect pervasive, culturally determined differences which would account for the differences in problem-solving style that were found in the present study. Further carefully planned cross cultural experimentation in this field is strongly urged.

Finally, the results of the multidimensional scaling analysis provide strong support for Guttman's[80)] (1965) theory of the structure of inter-relations among intelligence tests. In re-analyses of the original data of Thurstone[81)] (1938) and Thurstone and Thurstone[82)] (1941) using a smallest space analysis technique (Cf. Schlesinger and Guttman,[83)] 1969), Guttman[84)] (1965) was able to demonstrate a radex structure among the 21 tests comprising the battery. The tests separated into three segments corresponding to the three languages of communication used in the battery. These were verbal, numerical and pictorial. He also found that the tests requiring analytical skills were represented toward the centre of the configuration while tests measuring achievement skills were represented toward the periphery. In the present study tests of achievement were not included in the battery, but two segments of points, representing verbal and non-verbal (or pictorial) tests respectively, were clearly identified. Thus it may be concluded, that in addition to the formally identified mental processes required by a test's items, some variance in test scores is attributable to the medium in which the items are communicated. The implications of this finding should not be overlooked, when planning correlational experiments with tests of mental abilities.

8.5. Recommendations

The results would appear to support the use of the Thurstonian model of intellectual structure as a theoretical guide in developing a differential ability test battery for selection purposes in the CSIR. The Deductive Reasoning Test could be included with confidence in a battery of this nature. Suggested abilities to be measured by the battery would also include spatial reasoning, verbal meaning and induction. In constructing an instrument to measure the latter ability, the model proposed by Steyn and Verster[85] (1972) might prove a valuable point of departure.

REFERENCES

Order of Appearance

BALINSKY, B. (1941) An analysis of the mental factors of various age groups from nine to sixty, Genet. Psychol. Monogr., vol. 23, pp. 191 - 234. — 65

BLAKEY, R.J.A. (1941) A factor analysis of a non-verbal reasoning test. Educ. psychol. Measmt, vol. 1, pp. 187 - 198. — 20, 25, 49

BLOOM, B.S. and L.J. BRODER. (1950) Problem-solving processes of college students, Chicago, University of Chicago. — 66, 75

BOTZUM, W.A. (1951) A factorial study of the reasoning and closure factors. Psychometrika, vol. 16, pp. 361 - 386. — 22

BROWNE, M.W. (1969) Inter-group factor analysis. In: Grant, G.V., The organization of mental abilities of an african ethnic group in cultural transition. Unpublished Ph.D. Thesis, University of the Witwatersrand. — 53

CANISIA, SISTER M. (1962) Mathematical ability as related to reasoning and the use of symbols. Educ. psychol. Measmt, vol. 22, pp. 105 - 127. — 21

CLIFF, N. (1966) Orthogonal rotation to congruence. Psychometrika., vol. 31, pp. 33 - 42. — 55

COULTER, M.A. (1973) The effect of omissions on multiple choice tests. Psychol. Afr. (In Press.) — 39

FOUCHE, F.A. and N.E. ALBERTS. (1970) Preliminary Manual for the Senior Aptitude Tests. Pretoria, Human Sciences Research Council, Institute for Psychometric Research. — 24, 29

FRENCH, J.W., RUTH B. EKSTROM and L.A. PRICE. (1963) Manual for kit of reference tests for cognitive factors. New Jersey, Educational Testing Service. 9, 14, 15, 16, 23, 26, 27, 28, 30, 31, 32, 33, 35, 36, 42

FRENCH, J.W. (1965) The relationship of problem-solving styles to the factor composition of tests. Educ. psychol. Measmt, vol. 25, pp. 9 - 28. 68, 69, 71, 74, 78

GARDNER, R.W., P.S. HOLZMAN, G.S. KLEIN, LINTON HARRIET and D.R. SPENCE. (1959) Cognitive Control. Psychol. Issues, vol. 1, no. 4. 70, 76

GOWER, J.C. (1967) Multivariate analysis and multidimensional geometry. The Statistician, vol. 17, no. 1. 58

GRANT, G.V. (1969) The Organization of Mental Abilities of an African Ethnic Group in Cultural Transition. Unpublished Ph.D. Thesis. University of the Witwatersrand, Johannesburg. 54

GUILFORD, J.P. (1967) The nature of human intelligence. New York, McGraw-Hill. 8

GUTTMAN, L. (1965) The structure of interrelations among intelligence tests. In: Proceedings of the 1964 Invitational Conference on Testing Problems. New Jersey, Educational Testing Service. 80, 84

GUTTMAN, L. (1969) Integration of test design and analysis. In: Proceedings of the 1969 Invitational Conference on Testing Problems. New Jersey, Educational Testing Service. 59

JENRICH, R.I. and P.F. SAMPSON. (1966) Rotation for simple loadings. Psychometrika., vol. 31, pp. 313 - 323. 47

JÖRESKOG, K.G. (1963) Statistical estimation in factor analysis. Uppsala, Almquist and Wiksell. 48

KAISER, H.F. (1965) Psychometric approaches to factor analysis. In: Proceedings of the 1964 Invitational Conference on Testing Problems, pp. 37 - 45, Princeton, Educational Testing Service. 46

KAISER, H.F. (1970) A second generation little jiffy. Psychometrika, vol. 35, pp. 401 - 415. 45

KNEEBONE, G.T. (1963) Mathematical logic and the foundations of mathematics. London, van Nostrand. 11

LIENERT, G.A. and H.W. CROTT. (1964) Studies on the factor structure of intelligence in children, adolescents and adults. Vita Humana., vols. 3/4, pp. 147 - 163. 79

LINGOES, J.C. (1964) New computer developments in pattern analysis and nonmetric techniques. Paper presented at I.B.M. Symposium: Computers in Psychological Research. Blarium, Netherlands, pp. 24 - 27. 61

LUCAS, C.M. (1953) Analysis of the Relative Movement Test by a method of individual interviews. New Jersey, Educational Testing Service. 67

MEREDITH, W. (1964) Rotation to achieve factorial invariance. Psychometrika., vol. 29, pp. 187 - 206 (b). 52, 57

NASIONALE BUREAU VIR OPVOEDKUNDIGE EN MAATSKAPLIKE NAVORSING. (1958) Die Afrikaanse Woordtelling. Pretoria, Departement van Onderwys, Kuns en Wetenskap. 38

SAUPE, J.L. (1961) Some useful estimates of the Kuder-Richardson formula number 20 reliability coefficient. Educ. pschol. Measmt, vol. 21, pp. 63 - 71. 43

SCHLESINGER, I.M. and L. GUTTMAN. (1969) Smallest space analysis of intelligence and achievement tests. Psychol. Bull., vol. 71, no. 2, pp. 95 - 100. 60, 83

STEYN, D.W. and J.M. VERSTER. (1972) The development of a high level Symbol Grouping Test of determinative induction. Part One : Theoretical Considerations. CSIR Special Report no. Pers. 178. Johannesburg, SACSIR, NIPR. 10, 85

THOMSON, G.H. (1956) The factorial analysis of human ability. London, University of London. 56

THURSTONE, L.L. (1938) Primary Mental Abilities. Psychmtr. Monogr., no. 1. Chicago, University of Chicago. 1, 7, 34, 81

THURSTONE, L.L. (1940) An experimental study of simple structure. Psychometrika, vol. 5, pp. 153 - 158. 2

THURSTONE, L.L. and T.G. THURSTONE. (1941) Factorial studies of intelligence. Psychmtr. Monogr. no. 2, Chicago, University of Chicago. 6, 12, 13, 17, 82

THURSTONE, L.L. (1947) Multiple factor analysis : a development and expansion of "The Vectors of the Mind", Chicago, University of Chicago. 3, 44, 63, 64

VERSTER, J.M. (1970) The theoretical model for the construction of a test of deductive reasoning. Part One: Theoretical considerations. Johannesburg, SACSIR, NIPR. 4, 18, 19

VERSTER, J.M. (1972) The Measurement of Deductive Reasoning. Unpublished Masters thesis. University of South Africa, Pretoria. 5, 40, 41, 50, 51, 62

WITKIN, H.A., R.B. DYK, HANNA F. FATERSON, D.R. GOODENOUGH and S.A. KARP. (1962) Psychological differentiation : Studies of mental development. New York, John Wiley. 73, 77

WITKIN, H.A., HELEN B. LEWIS, M. HERTZMAN, KAREN MACHOVER, PEARL B. MEISSNER and S. WAPNER. (1954) Personality through perception. New York, Harpers. 72

WRIGHT, C.W. (1965) An English word count. Pretoria, Department of Education, Arts and Science. 37